Development of Local Program Theory: Using Theory-Oriented Evaluation to Make a Difference

KATHRYN A. SIELBECK-BOWEN
Bowen's Evaluation & Consulting Services, Athens, Pennsylvania

The aim of this article is to explicate the local program theory of an adolescent pregnancy and parenting program using a mix of methods. By bracketing these methods, local program theory is developed and subsequently used to distinguish between the validity of program theory and the validity of program implementation. Theory is then used as a tool to construct a model of how the program is supposed to work. This is followed by an assessment of the extent to which the program's theories are actually implemented in the field. The intent of using a mix of methods during this process is to gather an extensive, intensive, and comprehensive view of program case managers' thoughts about their work. Analysis is based on the evaluator's recognition of patterns of meaning that contribute to an enhanced contextual and theoretical understanding of the overall effectiveness of this program.

$\mathbf{P}$ablo Ruiz Picasso, one of the most influential artists of the twentieth century, challenged the traditional concept of an orderly, constructed, unified picture that mirrors the world. He presented the reality of forms in space from several angles of vision, introducing the idea that reality can be separate from appearance. In my evaluative work, I also seek to understand and know this reality, which may be different from what it seems to be. The world of social science is complex, contextually contingent, and mediated by individual interpretations of it. "A definitive account of the phenomenon of interest is not possible, although accounts based on extensive, intensive and comprehensive views can be better than others" (Smith 1997:78).

Local program theory development can be used to illuminate reality during the evaluative process and intensify our ability to meaningfully and legitimately know and understand the reality that exists within a social program context. Informed contributions can then be related to evidence of program effectiveness and program merit. This approach to evaluation tends to be more responsive to those who have a stake in a program, thus increasing the likelihood of producing evaluation results that inform, that are usable, and that strengthen programs. Evaluators who are interested in investigating

these varying perspectives on a phenomenon of interest by using a mix of methodologies will find this article particularly relevant.

The context for this discussion stems from an evaluation of a teen pregnancy and parenting program in central New York state. I will describe the development of program theory, testing the validity of program theory and the validity of program implementation by applying methodological bracketing of four methodologies. I will then assess the congruency between staff's theories, programmatic findings, and theories/outcomes of the external program planners.

OVERVIEW OF THE EVALUATION

During the evaluation of this teen pregnancy and parenting program, four distinct stages of development and evaluation of program theory were identified. During stage 1, focus was on the delivery of program services and outcomes as designed by the external program planners. Chen (1990) coined this particular program subtheory *normative treatment theory.* According to Chen, normative theory deals with what the structure of the program "should be, including such things as treatments, outcomes, and implementation processes that are related to the values of the program" (p. 43).

Stage 2 involved assisting program staff members in developing a statement of their program theory, which outlined the nature of the program as they interpreted it (staff members' local program theory). This program statement focused on staff members' beliefs about how the program worked in terms of the impact they believed their actions and interventions had on the behaviors of adolescents enrolled in the program. Despite the absence of immediate, observable outcomes, staff members had theories about which program effects they anticipated as a result of specific intervention strategies and assumed that certain outcomes were likely to occur from certain processes. These beliefs and theories were often gut feelings and tacit knowledge and were not always easily articulated.

To construct program theory based on these beliefs and assumptions of program staff, emphasis was placed on inductive strategies as opposed to logical deduction from a priori knowledge. This required an evaluation design that enlisted a mix of methodologies as well as maintaining responsiveness by actively involving stakeholders in the process of theory articulation (Greene 1993:14; 1994:533).

In stage 3, I conducted an investigation as to how the program was actually implemented and how program participants experienced the program. Then I assessed the congruency between staff members' program theory and the

actual implementation of the program. "This type of evaluation focuses on identifying the normative structure of the treatment, examining the actual treatment delivered in the field and assessing the congruency between the normative and implemented treatment" (Chen 1990:52; Chen and Rossi 1989:299).

Finally, in stage 4, the congruency between staff members' program theory and the theory of program planners was determined.

PROGRAM CONTEXT

The adolescent pregnancy and parenting program under study was established in 1983 in accordance with the New York state legislative mandate known as TASA (Teen Age Services Act). This piece of legislation mandated that all counties in the state of New York provide services to pregnant and parenting adolescents to maintain and strengthen family life and to attain or retain the capability for maximum self-support and personal independence. This evaluation took place in a county where a community-based program had been contracted to carry out this legislative mandate. TASA programs provided comprehensive case management services to each teenage participant and her immediate family (e.g., comprehensive needs assessment, goal plan development, service referral and coordination, continuous counseling, goal plan amending, comprehensive reassessment, and exit planning). Program participation was voluntary.

According to program records, 60%–70% of all referred pregnant and parenting teens agreed to join the program, and involvement with the program was typically for two years. The census was constant between 1996 and 1998, averaging 166 clients each year. Some of the program outcomes were positive: Compared with 1996 statistics, in 1997 fewer program participants were welfare dependent, homeless, abused drugs, and had experienced sexually transmitted infections, and more teens were delaying secondary pregnancy.

However, program statistics reflected several outcomes among participants that were not in line with their established program outcomes: (1) High school attendance decreased, from 49 in 1995 to 26 in 1998; (2) graduations from high school decreased, from 16 in 1995 to 9 in 1998; (3) teen pregnancy increased, from 69 in 1995 to 101 in 1998; (4) the number of parenting teens increased, from 87 in 1995 to 154 in 1998; (5) the number of teen mothers receiving child support from the fathers of their children decreased; (6) the number of teen mothers who improved their independent living skills decreased; and (7) the number of foster care placements increased during

1997 compared with 1996. These outcomes were unexpected and will be discussed later.

METHOD

The following discussion outlines the methods that were used during stages 1 and 2 to articulate the program theory as planned and subsequently explained by program staff. During stage 3, a snapshot of how the program was implemented by the staff will be provided, followed by an assessment of how the program was experienced by a sample of pregnant and/or parenting adolescents enrolled in the program. A discussion will follow describing the method used in stage 4 to identify parallels and disjoints between staff's local program theory and the normative theory of program planners.
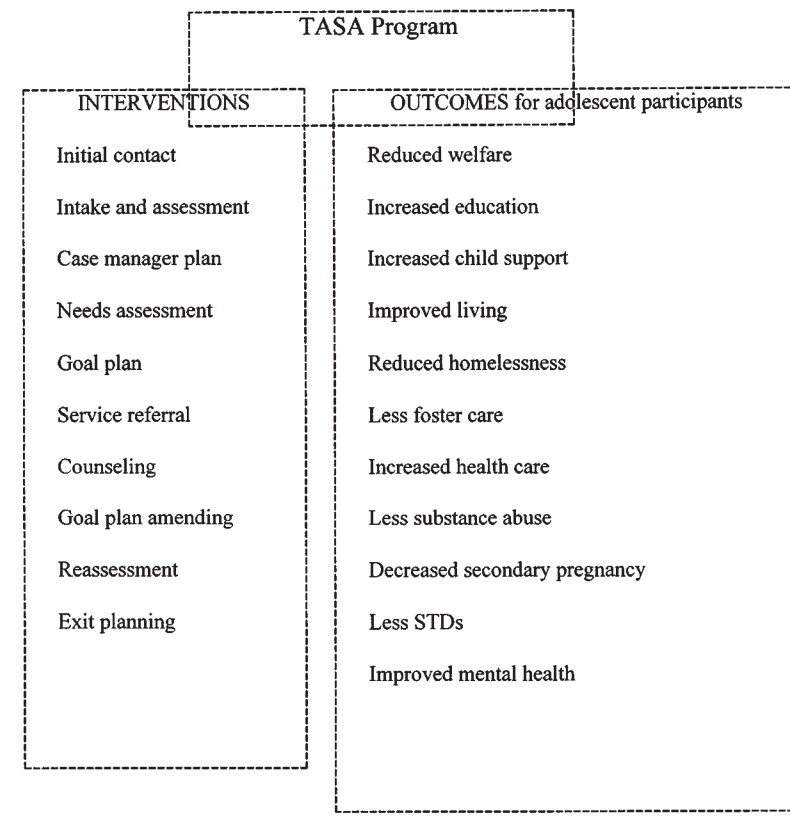
Document Review

Program records and correspondence over a three-year period (1995–1997) were reviewed to gain understanding of the program as it was planned. This review included organizational rules, regulations, financial and budget records, correspondence to program staff from state officials, and established goals and objectives. Annual reports and any other official or unofficial documents generated by or for the program were also examined in the initial days of this evaluative study.

These documents gave me information about how the program was supposed to be implemented that may not have been immediately observable during field observations. They also provided insight into how the program was envisioned by program planners and were a basic source of information about the program decisions and background, activities, and processes. I kept field notes on all these documents, placing special emphasis on identification of program planners' theories related to implementation of the program. Figure 1 depicts a model that reflects the program theory as identified during the review of documents. The logic of the planners was typical in that a list of interventions was expected to produce significant outcomes in a relatively short amount of time (Schorr 1997).

Staff Concept Mapping

In consideration of the multiple theories held by the program staff and my desire to be responsive, I conducted concept mapping with all five staff members. The aim of this evaluation was to establish a primary local program theory based on the latent theories that staff members held in their minds (Bick-

FIGURE 1
Logic Model: Program Theory as Conceived by Program Planners

TASA Program

INTERVENTIONS	OUTCOMES for adolescent participants
Initial contact	Reduced welfare
Intake and assessment	Increased education
Case manager plan	Increased child support
Needs assessment	Improved living
Goal plan	Reduced homelessness
Service referral	Less foster care
Counseling	Increased health care
Goal plan amending	Less substance abuse
Reassessment	Decreased secondary pregnancy
Exit planning	Less STDs
	Improved mental health

man 1987:5; Wholey 1987:77). These theories reflected how the implementation strategies (e.g., home visits, information, case management, transportation) were supposed to lead to the established outcomes (e.g., delayed secondary pregnancy, stable living conditions, higher level of education) (Chen 1998:15).

As Lipsey and Pollard (1989) concluded, concept mapping offers the evaluator a starting place for developing a program theory that is compatible with the perspectives and experiences of those who managed and delivered the program on a daily basis. This method was also an excellent way to estab-

lish a primary program theory statement where staff members contributed equally to the process and achieved consensus on the final statement of their local program theory.

The first step was to ask that staff members think about intervention strategies they believed led to program outcomes in addition to those integral to the overall success of the program. Staff members were then asked to brainstorm statements that described these important aspects of the program. One hundred twenty-two different statements were generated and entered into a Concept Systems Inc. (1998) program (Trochim 1989). Each statement was copied onto business-sized cards.

Staff members were given a stack of 122 cards and asked to sort the statements on the cards into categories that "made sense to them." Then staff members were asked to rate each statement in order of importance, with 1 being the least important intervention strategy and 5 the most important. To make the results more robust, this sorting and rating process was conducted two different times and entered into the Concept Systems software.
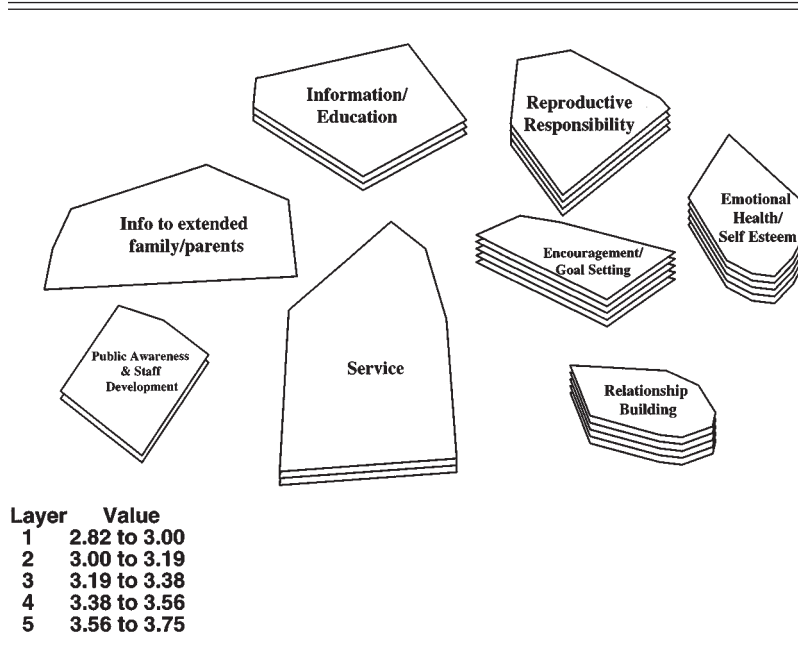
The software conducted a multidimensional scaling (MDS) analysis that locates the 122 statements as points on a map. Statements placed together by more respondents are considered more similar to one another than are statements placed together by fewer respondents. Statements that are close to each other on MDS are considered conceptually similar (because they were sorted together more frequently by program staff members), whereas statements that are far from each other on the map are considered conceptually dissimilar. (For more on MDS and concept mapping, see Trochim 1989.)

Cluster analysis is used to interpret the MDS output. An eight-cluster solution, shown in Figure 2, was produced. I was able to sit down with the program staff and discuss the meaning of the clusters of statements produced by the analysis and determined subjectively the optimum number of clusters. This qualitative exercise took full advantage of staff members' knowledge and experience in partitioning the statements according to locally meaningful categories.

Analysis of Cluster Rating Map

Clusters in the southeastern section of the map were rated as the most crucial to program success and theoretically were the primary focus of intervention strategies. These two primary clusters contained statements that described program activities related to relationship building and encouragement and/or goal setting (average rating 3.75/5.0). Emotional health and/or self-esteem focused interventions were centrally located in the east and represented another theoretically crucial aspect of the program (average rating

FIGURE 2
Cluster Rating Map

Layer   Value
1   2.82 to 3.00
2   3.00 to 3.19
3   3.19 to 3.38
4   3.38 to 3.56
5   3.56 to 3.75

3.57/5.0). Finally, reproductive responsibility (average rating 3.39/5.0), located in the northeastern section of the map, also represented a theoretically crucial aspect of the program.

The eastern section of the map focused on the adolescent client, with particular emphasis on building trust, establishing a relationship and rapport, bolstering self-esteem and self-identity, and supporting and encouraging thoughtful decisions related to the future in general, and sexual activity/contraception in particular.

Clusters located in the central section of the map, education and information as well as services, were also important aspects of program staff members' treatment theories and were rated only slightly lower than those theorized as being the most crucial (3.32–3.35 vs. 3.39–3.75). These clusters were directed more toward theories related to the tangible services that the Teen Pregnancy Program provided, such as the mentoring program, equipment and clothing, transportation, program newsletter, and availability for walk-in appointments (crisis intervention) in addition to information about

accessing outside community resources and advocating for the teen when necessary.

The cluster titled "public awareness and staff development" (average rating 3.32), located in the southwest corner of the map, relates to staff members' theories about the impact of public awareness and attending to their own needs on successful program implementation. Finally, to the west and theoretically the least crucial to successful implementation of the program are the services directed at health and information to extended family (average rating 2.82). Clusters with the highest ratings (3.32 or greater) were carried forward and used as key concepts to be probed more thoroughly during staff interviews.
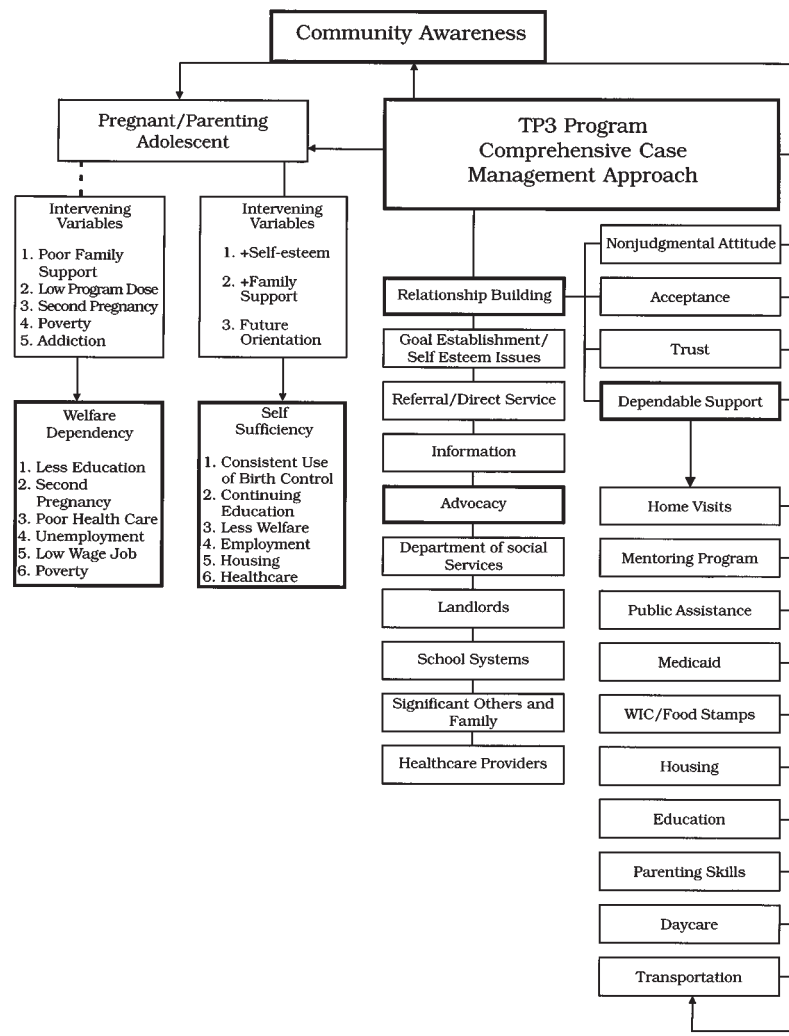
One-on-One Interviews

The purpose of the interviews was to amplify and clarify staff members' local program theories on an individual basis while decreasing the threat of social desirability that might have affected the validity of results during the total group concept mapping session. Each interview was transcribed word for word, entered into Ethnograph (4.0), and coded according to the key concepts of the program as identified during concept mapping. Staff members were given opportunities during the interview to add to the theory base established during concept mapping and to dispute concept mapping results. Upon completion of staff interviews, I constructed a case-ordered descriptive meta-matrix that illustrated the bracketed findings from the two methodologies utilized. This matrix was used to assess the fit between concept-mapping results and interview responses across staff members (Miles and Huberman 1994).

The analysis was intended to capture meaning and action in contextually sensitive snapshots and narratives that described program staff members' experiences in delivering the program and beliefs about their work and the impact of that work (Greene 1990:228). Precision of measurement, interobserver reliability, and generalizability were not as important to me as the preciseness of concept mapping and intensiveness of one-on-one interviews.

Based on concept mapping and individual one-on-one interviews with staff, I constructed a logic model that reflected the results (see Figure 3) (Donaldson 1993:195). I showed this model to staff members and requested their input. After minor adjustments, the staff reached consensus and adopted the local program theory model as being reflective of "the mechanisms that mediated between the delivery and receipt of the program and the emergence of the outcomes of interest" (Weiss 1998:56). Program staff members' outcomes of interest were not necessarily always directly in line with planners'

FIGURE 3
Logic Model: Staff Members' Local Program Theory

outcomes, however. This will be discussed in some detail in the section on findings.

STAGE 3: ASSESSMENT OF THE PROGRAM
AS IMPLEMENTED

During this stage of the evaluation, I conducted an assessment of the congruency (or lack of congruency) between staff members' local program theory and the program as implemented. I used four methods to assess congruency between theory and practice: chart review, observations, concept mapping, and interviews with teens.

Step 1: Chart Review

An initial methodological step involved reviewing program participants' charts. Detailed charts were maintained on every program participant and included an intake form (general information related to age, address, phone number, state of pregnancy/parenting, economic situation, and level of education), past medical history, a statement of goals, and progress notes.

A purposive random-sampling technique (Rubin and Babbie 1993:255) was used to select typical-case adolescents enrolled in the program. The program director was asked to use her judgment and handpick clients who she thought represented the typical adolescent enrolled. Selection criteria were established that included adolescents between the ages of fifteen and eighteen who had been in the program for at least one year. These criteria were established so that the program participant would have been enrolled long enough to have gained an understanding of as well as experienced the implementation process.

A list of 100 names was generated. I selected every fifth name from this list and used this process to decrease the potential for bias during the selection process. I didn't have any knowledge of each teen's level of interaction with the program in advance of selection nor did I know her personal/health/birth history. The rationale for this selection process was to reduce questions as to why certain teens were selected whereas others were not (realizing that this sample, although typical, was not representative).

I systematically selected and reviewed twenty-five charts. I took field notes on individual charts, paying special attention to documents that reflected the crucial aspects of the program identified in staff members' local program theory. Additionally, a checklist matrix was completed for each chart reviewed. This checklist identified the level to which staff members actively implemented their theories based on documented evidence in client charts. For example, did staff document evidence of relationship building to a high degree (more than four entries), moderate degree (two to four entries), or low degree (one or fewer entries), and/or did staff document evidence of

being accepting and nonjudgmental to a high, moderate, or low degree? These documents gave me a behind-the-scenes look at the program processes in terms of interactions with several adolescent participants.

Client records were written in an informal, conversational style. Notes were extremely detailed process records and clearly outlined each client's encounter with her case manager, both casual and scheduled. The charts served as diaries of chronological events and as a ledger of sorts that could be used to validate the services that were administered; they also showed the level of program implementation each teen received. These client records were the only source of routine tracking and included descriptions of program participants' reactions to the strategies as well. Records varied, however, in detail and thoroughness across case managers.

Step 2: Home Visit Observations

I conducted observations of home visits to learn and understand more fully about the program beyond what staff and participants stated. I selected a convenience sample of fifteen home visit observations. These teens may or may not have had their chart reviewed. Because of limited time and resources, observations were based on convenience, where the objective was to observe each case manager during a home visit at least twice. The observational strategy I used was that of an onlooker or nonparticipant, and I maintained a low profile during the visit. I observed home visits with case managers in order to see the interaction between program staff and program participants and to assess the level of program implementation. This first-hand information helped me identify the level of direct program implementation. A list of sensitizing concepts (based on the key aspects identified in the local program theory) was constructed before these observations to focus, organize, and guide the observational fieldwork (Patton 1997:216).

Results from the observations were methodologically bracketed to the results from the chart review to assess congruency of what was written and what was observed. Overall, many of the interactions, activities, and nuances of the home visits were not documented in the client's chart.

Step 3: Teen Concept Mapping

A convenience sample of adolescents was selected to participate in a concept-mapping session. Staff members were asked to assess interest among those clients in their caseload between the ages of fifteen and nineteen years who had been in the program for at least six months.

The concept-mapping session required two hours of brainstorming, where seven participants were asked to make statements that described the program

treatments they found to be the most important (generally and then specifically related to contraceptive use, education, and health care practices). Teens followed the same procedures as described above. I used the completed concept map as an initial means of understanding the program as experienced by the adolescents enrolled.
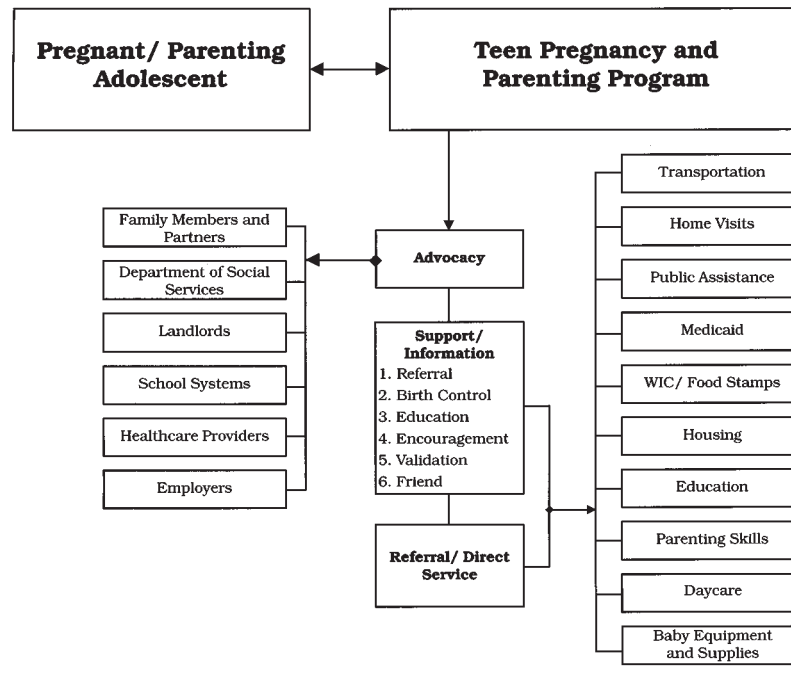
Step 4: Teen Interviews

Typical case sampling was used to select twenty-five adolescents for one-on-one interviews. These young women were between thirteen and eighteen years of age, had been enrolled in the program for at least one year, and were in various stages of pregnancy/parenting. Questions for the interview guide emerged as a result of an iterative process that occurred during chart reviews, observations, and concept mapping. Interviews were transcribed and entered into Ethnograph. Key concepts for the interviews were drawn from concept-mapping results and followed the same interviewing procedure as described for staff. A logic model was constructed that depicted teens' interpretation of the program based on information collected during concept mapping and one-on-one interviews (see Figure 4). Overall, teens recognized the support and relationships, as well as the tangible items, provided by program staff. Concept-mapping results were again bracketed with interview results in order to arrive at this logic model.

DATA ANALYSIS PROCEDURE: METHODOLOGICAL BRACKETING

I had two major assumptions. First, congruencies between the local program theory as identified via concept mapping and one-on-one interviews would produce a strong statement reflecting staff members' theories that could be carried forward during the evaluation. On the other hand, inconsistencies between the two sets of data would make stating clear program logic problematic. This lack of clarity would be considered a program weakness in that implementation of the program in general could be affected directly.

Second, congruencies between the local program theory of staff and empirical data collected during stage 3 of the evaluation (teen concept map, observations, chart review, and interviews) would indicate program strength. Disparity between staff members' local program theory and empirical data would be indicative of program weakness and point to potential problems in fully implementing the program (Chen and Rossi 1989:300).

FIGURE 4
Logic Model: Program as Experienced

Methodological bracketing aided in making connections between descriptive information, program process, and evidence of program impact. Concept mapping, observations of home visits, chart reviews, review of program documents, and personal interviews with staff and participants all worked as data sources to make these connections.

The initial data analysis procedure was that of bracketing concept-mapping results with one-on-one interview findings in an iterative manner. Bracketing was an alternative to simple convergence where nonidentical results of the different methods were used to specify a range or bracket that presumably contained the correct answer (Reichardt and Cook 1979). The benefits of bracketing during this program evaluation enriched and deepened my understanding of the program environment while I sought convergence, corroboration, and correspondence of results across the different method

types (Mark and Shotland 1987:95). Each method was considered equally sensitive and weighted equally. The use of methodological bracketing maximized the strengths and minimized the weaknesses of each individual methodological approach while strengthening the evaluation findings and contributing to theory and knowledge development.

The process of analyzing these data included repeated readings of the data as a whole and then arriving inductively and intuitively at a set of credible theories. These theories were presented to the staff to assess whether they truly reflected how program staff members conceptualized their work. I then established the warrant for each theory by bracketing the findings from one methodology with the findings of another, assembled the confirming evidence from the data record, searched for disconfirming evidence, weighed the evidence one way or the other, and then constructed a logic model to visually depict the findings.

For example, staff ratings during concept mapping indicated that relationship building, trust, respect, and nonjudgmental attitudes were the most important aspects of the program. They were validated as being the most important when bracketed with the data collected during one-on-one interviews. The bracket of findings from concept mapping and staff interviews was very narrow, essentially converging on this one aspect of the program theory as articulated by the staff:

> I think part of that, too, is before you can do anything with them, you have to establish some sort of relationship. (Staff interview 1997)'
>
> We develop a relationship and it's in a nonjudgmental way, they call the shots and that is important. (Staff interview 1997)

This was further corroborated during stage 3 of the evaluation when teens were interviewed one on one, thus validating that this aspect of the local program theory was actively being implemented in the field:

> She is very easy to talk to, she is nonjudgmental, you just like to have her there for somebody to talk to about things that you even don't feel comfortable talking to your family about. (Teen interview 1998)
>
> They help me figure out what I needed to do and how to go about it. They're supportive and understanding, they do their best to help you, they're nonjudgmental, so that helps. (Teen interview 1998)

Another example of the merits of methodological bracketing occurred during assessment of the level to which the program was actually implemented. Field notes from the review of patient charts were bracketed with observations of home visits to make connections between what staff theo-

rized as important implementation strategies, what they documented in client records, and which strategies were actually being implemented, as well as to what level.

I also reviewed client records to counterbalance the reactive influence of my presence during observations of home visits. The methods selected had dissimilar biases and, therefore, were anticipated to produce less systematic effects of participant- and evaluator-based errors, which lead to problems with the credibility and utility of the evaluation.

Two examples follow that highlight disjoints between findings from different data sources (creating very wide brackets). (1) While the client chart may have included a detailed, thorough goal plan that included plans to enroll in a GED program, find employment, or move into a more suitable home (as identified in the local program theory), observations during home visits revealed no references to these goals as written; rather, emphasis was placed on present-day concerns such as informational needs, day care, and transportation. Documentation for the visit usually reflected only a cursory entry (or no entry at all) related to the information shared. Observations revealed detailed discussions and demonstrations of, for example, birth control information, various options for day care, and offers to provide transportation with engaging, warm interchanges between the case manager and her adolescent client. (2) While concept-mapping results indicated that assisting teens to become more sexually responsible was one of the most important aspects of the program, observations revealed little emphasis on responsibility. Instead, focus was placed on information sharing with regard to birth control options.

Through this process of making connections and achieving clarity via methodological bracketing, I came to see how well the program theory described what actually was happening in the program. I then ascertained the viability of the program theory by comparing the expectations articulated in the theory with empirical data to see how well they connected to one another (Weiss 1998). Table 1 outlines examples of comparisons that were drawn during this data analysis procedure.

FINDINGS/EVIDENCE OF SUCCESS BASED ON STAFF MEMBERS' LOCAL PROGRAM THEORY

Logic models were used as pictorial depictions of the findings during presentations to the program staff. Comparisons between models illuminated congruency and lack of congruency between staff members' theories and what was actually implemented.

TABLE 1
Cross-Comparison of Evaluation Results

Crucial aspects of the Teen Pregnancy Program as established by external program planners	Crucial aspects of the Teen Pregnancy Program's program theory as identified by concept mapping	Crucial aspects of the Teen Pregnancy Program's local program theory as identified by one-on-one interviews	Crucial aspects of the Teen Pregnancy Program as identified by teen participants	Comments
(note: no goals aligned with relationship building)	Relationship establishment received one of the highest ratings (3.75).	Staff believed that the establishment of relationships built upon trust, respect, and nonjudgmental attitudes was foundational for the success of all the intervention strategies.	Teens clearly recognized this emphasis on relationship building. Ninety percent of teens interviewed provided evidence during their interview of a relationship with their case manager as well as a high level of trust ($n = 25$).	Client records were filled with entries reflecting informal process outcomes that support this particular aspect of local program theory. This aspect was not measured formally, however. Concept-mapping and one-on-one interview results converged on this
(1) Progress toward reduced welfare dependence, (2) increased level of education, (3) increased child support, (4) improved independent living skills, (5) reduced homelessness or housing vulnerability, (6) prevention of foster care placement,	Also, staff concept-mapping result identified goal establishment as one of the most important aspects of the program with a similar rating of 3.75.	Overall consensus during interviews was that the goals as established by planners were unrealistic and difficult to achieve. Staff recognized small achievements (not necessarily expecting program goal achievement), realizing that some teens are under incredible stress and	Teen concept-mapping results *did not* include goal establishment as an important aspect of the program. Personal goals were established by the adolescent participant (often not directly reflective of the outcomes as established by program planners). Examples of teen goals are to obtain a car and	aspect, creating a very narrow bracket. While "small steps" were not measured beyond intermittent and inconsistent documentation in client records, program goals were checked off as achieved or carried forward in the goal plan every six months (this was part of the mandatory

find safe day care for their children.

Teens rarely remembered short-term or long-term goals they had identified during their intake inter-views (96%). Most did not connect positive strides directly with their participation in the program. The program was viewed as a stable, dependable source of advocacy and support, as well as a source of tangible items (rated the highest among concept-mapping participants) to meet their needs such as diapers, clothing, transportation, and infant formula. The relationships adolescents shared with their case managers were perceived as most important according to interviews.

reporting for funders/planners). These data were maintained and sent forward to the state in the form of an annual report, and were used as the primary yardstick of program success or mediocrity. Bracketing the results of document review (program as planned) and results of staff members' interpretation/implementation of the program produced the widest bracket, indicating very little convergence. This indicated a significant departure from what staff theorized as an important aspect of the program and what was actually implemented in the field. Additionally, results from teen concept mapping and interviews further corroborate this departure in that there was little transfer of this focal point to them.

(7) increased healthy pregnancy outcomes, (8) increased effective use of health care, 9) reduced substance abuse, (10) decreased rate of secondary pregnancy and sexually transmissible infections, and (11) improved mental health.

have long histories of poverty, addiction, instability, and chaos. Goals established by planners were considered ideal by program staff, but the reality of day-to-day implementation of the program and assisting teens in crisis precluded the feasibility of achieving most of these goals to a high level in Teen Pregnancy Program strides directly.

Because this evaluation was responsive to program staff members themselves, it was discovered that staff members' primary objective was not goal achievement and behavioral change as established by program planners. Rather, staff members placed their energy into building relationships with the adolescents enrolled. As a result, most of the adolescents interviewed and observed recognized a trusting/nonjudgmental relationship with their case manager as opposed to being able to identify specific ways their behaviors had been changed or how their decisions were influenced directly.

This focus on relationships was of extraordinary importance among the teens in the program (according to staff interviews) because many did not routinely experience trusting relationships with adults. Their parents frequently had not been able to foster trusting relationships because of poverty; under- or unemployment, drug or alcohol addiction, and lack of education had hindered their ability to provide environments conducive to nurturing relationships. Furthermore, many of these families were not always capable of providing the basic positive developmental foundation necessary for the healthy psychological growth of the adolescent. This type of background produced adolescents who had low self-esteem and low self-efficacy in general (Moore 1995:150). Because this program targeted these vital aspects and provided external support in these critical areas, there was potential for long-term positive effects that were not immediately measurable at the time of this evaluation (over a six-month period).

Observations and chart reviews revealed that many of the pregnant and parenting adolescents enrolled in the program had chaotic and complex lives. According to the case managers, they lacked the security, self-efficacy, self-esteem, and self-identity necessary to make informed decisions or serious plans about their future. The program as planned established primary outcomes such as self-sufficiency, welfare independence, stable housing conditions, higher education, and employment. While ideal, these outcomes were rather unrealistic and bounded far beyond the point at which the typical teen participant existed daily. Staff members recognized this and focused on building the teen personally with the hope that ultimately the young mother would have the personal resources to attain some of the goals as established by the program planners at some point in the future.

Case managers believed that if a trusting relationship was established, positive outcomes would follow. Although relationship building did not occur with each individual teen, case managers focused unconditionally on this result. Of those teens who were open to a relationship with their case manager, positive outcomes were often observed (continuation in school, delayed secondary pregnancy, stable housing).

Small positive steps were recognized by the case managers and gave the teens positive reinforcement not only for major achievements such as continuing education or delaying a secondary pregnancy but also for small achievements like following through with scheduled appointments and maintaining a safe environment for her child. This recognition of small steps helped the case manager establish a relationship with the teen and ultimately led to evidence of enhanced self-esteem and self-identify (this was subjective evidence provided by case managers, since no formal measurement of self-esteem was conducted).

Staff attended to each program participant's self-esteem and self-efficacy by providing tangible skills, a progressive increase in responsibility, and opportunities to give to others. In some circles, these contributions to program participants would be considered an indicator of successful program implementation (Schorr 1997). This success was achieved specifically by (1) educating the adolescent about contraception, sexually transmitted diseases, community services and resources, childbearing, childbirth, and parenting; (2) modeling positive behaviors and then expecting the teen to reflect these; and (3) offering teens opportunities to make a meaningful contribution by participating on peer panels that informed other youth in the community about issues of teen pregnancy and parenting. Case managers viewed their program participants as having complex constellations of strengths and weaknesses. They attempted to build on the strengths while targeting the weaknesses. These strong theories affected implementation of the program as seen in the documentation of activities, observations during home visits, and validation during interviews with adolescent participants themselves.

FINDINGS: MODERATE SUCCESS BASED ON PROGRAM PLANNERS' THEORIES

During the day-to-day operation of this teen pregnancy and parenting program, there were many occasions when informal counseling, consultation, information sharing, instruction giving, advice sharing, and referrals for needed services occurred. These were not planned, and neither standardized intervention strategies nor the impact of these serendipitous interactions was measured. This lack of standardization, variability in program dose, and lack of formal means of measuring short-term indicators of movement toward long-term outcomes resulted in only moderate evidence of program success overall.

For example, case managers might have had a well-laid-out plan to view a birth video with the expectant mother and follow that with a discussion and

questions, the intent being to work toward the goal of educating the teen and promoting sexual responsibility. In reality, the birth video may never be viewed and discussions related to responsibility may never occur because each time the case manager attempts to implement her plan, a crisis or more pressing need emerges. It was not uncommon for case managers to arrive at the teen's home to find that the landlord had just delivered an eviction notice, the house was in turmoil because of numerous visiting family members, the teen had many basic needs, or an abusive boyfriend had returned. Case managers were observed competently responding to these spontaneous events; however, there was no tracking or monitoring of these activities in place.

This evaluation produced evidence that suggested that the establishment of context-free program goals by external planners may have been a recipe for the illusion of program failure in this situation. This illusion might very well lead to the reality of program discontinuation despite the positive impacts the program is making on adolescents' lives. This program, like many others, had a paucity of indicator data that could accurately document the effects of each intentional staff intervention. Sorely missing was the tracking and measurement of interim milestones that could show reliably that staff members' focus on relationship building did indeed have a positive effect on program participants, assisting them to reach program goals as established at some point in the future.

An interesting situation arose in that the people who planned the program did not share the same logic as people implementing the program. Because external planners were context free, their plans (while well meaning) for the program reflected generic preventive or ameliorative strategies drawn from the literature and/or personal experiences.

Furthermore, the political and personal agendas of politicians/decision makers may have driven the establishment of outcomes for the program. Moreover, the planners' perspective was based primarily on a technical rationality perspective. The emphasis is on techniques and processes that are not only context free but also implementer free. This approach assumes that case managers implement the program in identical, consistent ways and that anyone with the qualification necessary for the job can provide or deliver the program services equally to each participant.

Planners did not anticipate that implementation of this program would be so deeply embedded in the relationships that were formed during the process of service delivery, or that these relationships would be so intricately intertwined with the personal and interpersonal characteristics, knowledge, and skill of each individual staff member (Friedman 2000).

In addition, the goals as established by program planners, such as self-sufficiency, continuing education, delayed secondary pregnancy, stable

housing, and improved health care, assume that pregnant and parenting adolescents live in environments and have personal characteristics that are conducive to significant influence by program staff in a rather short period of time. This assumption was clearly not substantiated in this evaluation.

A significant portion of what case managers actually did (their program activities and implementation strategies) during home visits, on the phone, in the office, and in the community were not standardized, measurable activities. Daily, they were responding to contingencies. Time, patience, support, and advocacy were required over time in order for relationships to be built. Once a trusting relationship was established, the case manager could then move forward, working more intensely on behavioral change. Time required for establishing relationships was extremely variable and depended on personal characteristics, the environment, and the level of contact the teen had with the program. This evaluation validated the claim that a much broader level of inquiry is needed to understand the dynamic conditions that exist in human service programs such as the one described.

A QUESTION OF UNINTENDED CONSEQUENCES?

As suggested above, program outcomes between 1995 and 1997 reflected only moderate impact related to program participation. Most surprising was the increase rather than decrease in adolescent participants who were pregnant and parenting. Equally surprising was the increase in high school dropout rates among program participants and how few chose to continue their education after high school. The question that has to be asked is that of unintended consequences and the possibility of actually doing harm as opposed to good.

The plausible explanation offered by the program director for the poor outcomes was related to the impact she believed that welfare reform was having on educational attainment of participants and the unrealistic expectation that teens should enter the workforce by the age of eighteen. This new regulation pushes teens into minimum-wage-paying positions that provide no benefits and little hope of escaping the downward spiral of poverty over time.

Alternatively, the questions that have to be considered are: (1) What are the unintended consequences of this program? and (2) Is the program doing harm as opposed to good? Some members of the community agreed with this alternative hypothesis of doing harm and complained that this program for pregnant and parenting adolescents made life "too easy" for the youngster, in that case managers took responsibility and made excuses for the teen. The provision of transportation, food, clothing, shelter, advocacy, and guidance

often led to a "them" and "us" scenario, where other community service providers were viewed as not understanding the plight of the adolescent parent and being indifferent, uncaring, and rude by program staff. Program staff members frequently found themselves in adversarial positions with community service providers, particularly those from the Department of Social Services, health care agencies, and the educational system.

Overall focus seemed to be placed on the process of program implementation, where the means and ends became confused and the focus on what actually happened to the adolescent over the long term was lost. Perhaps because this program purposely did not focus on goal achievement, goals as established by program planners were not fully achieved by many of the participants.

CLOSING COMMENTS

In this adolescent pregnancy and parenting program situation, where broad, complex, and interactive interventions were aimed at multiple outcomes, many of the influences that affected the overall program processes and outcomes did not stand still long enough to permit certainty about what caused the observed behaviors in adolescent participants. However, the incorporation of local program theory development, especially within an interpretivist philosophical inquiry framework using a mix of methods, served to illuminate multiple perspectives related to key program issues in this particular context and began to disentangle some of the complexities of context and lived experiences.

This theory-oriented evaluation identified a strong, defensible, and perhaps uncommon program logic. Program staff disclaimed possible impact on longer term behaviors in favor of planting the seeds for teens to nurture and actualize on their own. For program staff, the establishment of a trusting, nonjudgmental relationship with the teen was the major goal. If staff members could do this successfully, and if other environmental and personal conditions were favorable, teens could gain sufficient efficacy and confidence to begin to build their own independent, productive lives.

This adolescent pregnancy and parenting program logic focused on relationship building and demanded time and intense legwork on the part of program staff. Unfortunately, in the process of building these relationships and dealing with contingencies, staff members sometimes lost sight of long-term program goals. Without the staff's knowledge, the process of program delivery had overtaken concerns about and focus on the end product. Because teens were typically enrolled in the program for only two years, few of the

long-term goals were achieved. Highlighting the disparities between program planners' logic and the logic driving the program during the process of evaluation made the need for tracking and monitoring interim milestones much more evident to program staff.

My recommendations include highlighting this lack of focus on program goals and offering encouragement to be more intentional in keeping the long-term goals as a central focal point (to include tracking and monitoring outcomes over time). I also recommend that additional measures of interim milestone achievement are essential to show that program interventions are on track toward long-term goals and to increase accountability and focus on those goals. A more formal way of measuring self-esteem and self-efficacy is also recommended, with a measure to be administered at the time of the intake interview and another when the adolescent's case is closed.

Finally, I discuss the possible negative impact of focusing so intently on unconditional relationship building to include decreased accountability, experience, and responsibility for making good choices among adolescent participants, as well as encouraging a level of dependency. Staff members were encouraged to assess each case carefully and offer more opportunities in which teens are given responsibility, and were held accountable as opposed to being provided with advocacy and unconditional support, regardless of the circumstance. Again, staff members were encouraged to monitor and track these decisions and outcomes observed as a result.

USE OF FINDINGS

After this evaluation was completed, major staff changes occurred within the program. The program director and two case managers took other positions. Within one month, more than one-half of the staff had to be replaced. As of the writing of this article, evaluation recommendations have not yet been incorporated.

REFERENCES

Bickman, L. 1987. The functions of program theory. *New Directions for Program Evaluation: Using Program Theory in Evaluation* 33:5–18.

Chen, H. T. 1990. *Theory-driven evaluations*. Newbury Park, CA: Sage.

———. 1998. Theory-driven evaluations. In *Advances in educational productivity*. Vol. 7, *Evaluation research for educational productivity*, edited by A. J. Reynolds and H. J. Walberg, 15–34. Greenwich, CT: JAI Press.

Chen, H. T., and P. H. Rossi. 1989. Issues in the theory-driven perspective. *Evaluation and Program Planning* 12:299–306.

Concept Systems Inc. 1998. *The concept system, ideas in action, version 1.75*. Ithaca, NY: William Trochim.

Donaldson, S. I. 1993. Testing the generalizability of intervening mechanism theories: Understanding the effects of adolescent drug use prevention interventions. *Journal of Behavioral Medicine* 17:195–214.

Friedman, V. 2000. Action science: Creating communities of inquiry in committees of practice. In *Handbook of action research*, edited by H. Bradbury and P. Reason, 120–39. Thousand Oaks, CA: Sage.

Greene, J. C. 1990. Three views on the nature and role of knowledge in social science. In *The paradigm dialog*, edited by E. G. Guba, 227–45. Newbury Park, CA: Sage.

———. 1993. The role of theory in qualitative program evaluation. In *Theory and concepts in qualitative research*, edited by D. J. Flinders and G. E. Mills, 14–15. New York: Teachers College Press.

———. 1994. Qualitative program evaluation: Practice and promise. In *Handbook of qualitative research*, edited by N. K. Denzin and Y. S. Lincoln, 533. Thousand Oaks, CA: Sage.

Lipsey, M. W., and J. A. Pollard. 1989. Driving toward theory in program evaluation: More models to choose from. *Evaluation and Program Planning* 12:317–28.

Mark, M., and R. Shotland. 1987. Alternative models for the use of multiple methods. *New Directions for Program Evaluation* 35:95–99.

Miles, M. B., and A. M. Huberman. 1994. *An expanded sourcebook: Qualitative data analysis*. 2d ed. Newbury Park, CA: Sage.

Moore, K. 1995. *Adolescent pregnancy prevention programs: Interventions and evaluations*. Washington, DC: Child Trends, Inc.

Patton, M. Q. 1997. *Utilization focused evaluation*. 3d ed. Thousand Oaks, CA: Sage.

Reichardt, C. S., and T. D. Cook. 1979. Beyond qualitative versus quantitative methods. In *Qualitative and quantitative methods in evaluation research*, edited by T. D. Cook and C. S. Reichardt, 7–32. Beverly Hills, CA: Sage.

Rubin, A., and E. Babbie. 1993. *Research methods for social work*. 2d ed. Pacific Grove, CA: Brooks/Cole.

Schorr, L. B. 1997. *Common purpose*. New York: Anchor Books Doubleday.

Smith, M. L. 1997. Mixing and matching: Methods and models. In *Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms*. New Directions for Evaluation, no. 74, edited by J. C. Green and V. J. Caracelli, 78–85. San Francisco: Jossey-Bass.

Trochim, W.M.K. 1989. An introduction to concept mapping for planning and evaluation. *Evaluation and Program Planning* 12:1–16.

Weiss, C. H. 1998. *Evaluation*. 2d ed. Upper Saddle River, NJ: Prentice Hall.

Wholey, J. S. 1987. Evaluability assessment: Developing program theory. *New Directions for Program Evaluation: Using Program Theory in Evaluation* 33:77–92.

KATHRYN A. SIELBECK-BOWEN is an independent evaluation consultant who owns and operates her own evaluation firm in Athens, Pennsylvania. She specializes in conducting evaluations of programs for children and youth using multiple methods and concept mapping.