INVESTIGATING THE VALIDITY OF A MATURITY MODEL FOR STAGING ORGANIZATIONAL CAPACITY OF NONPROFIT COMMUNITY ORGANIZATIONS

by

Russell G. Schuh

BA, Wichita State University, 1970

Submitted to the Graduate Faculty of

School of Education in partial fulfillment

of the requirements for the degree of

Doctor of Education

University of Pittsburgh

2004

UNIVERSITY OF PITTSBURGH

SCHOOL OF EDUCATION

This dissertation was presented

by

Russell G. Schuh

It was defended on

April 27, 2004

and approved by

Dr. Louis Pingel

Dr. Mary Kay Stein

Dr. Laura C. Leviton

Dr. R. Tony Eichelberger
Dissertation Director

INVESTIGATING THE VALIDITY OF A MATURITY MODEL FOR STAGING ORGANIZATIONAL CAPACITY OF NONPROFIT COMMUNITY ORGANIZATIONS

Russell G. Schuh, EdD

University of Pittsburgh, 2004

The purpose of this study was to investigate the validity of a maturity model for staging the capacity of nonprofit community organizations to implement or sustain social interventions or demonstration programs. During the period 1999 to 2001, a five-stage maturity model was synthesized using knowledge and experience derived from diverse fields, and applied to small nonprofit organizations in an iterative process of (a) application, (b) analysis, (c) revision, and (d) re-application. Two raters with organizational experience, in addition to the developer, were trained to use the instrument. They participated in all phases of the formative development of the maturity model. The resulting estimation of the instrument's validity is based on convergent results of four analyses; (a) content validation by comparison of the new instrument with existing instruments assessing capacity to determine the extent to which it assesses important dimensions of organizational capacity, (b) construct validation by comparison of an early version of the model with a later version to assess its evolution, (c) estimates of interrater reliability among three raters, and (d) construct validation through feedback from agency staff and governing board members and feedback from staff involved in funding those agencies. The results of these analyses are mixed, not establishing statistical conclusion validity, but showing promise for the instrument in its formative stage of development.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

This study is the culmination of five years of development during which many organizations and individuals provided invaluable assistance and support. The study reported here is based on work performed for the Small Agency Building Initiative (SABI) of the Robert Wood Johnson Foundation and related grants supporting the development of the Staging Organizational Capacity (SOC) instrument. The nineteen SABI grantees have all been profiled using the SOC. Several other organizations, not related to SABI, volunteered to have the SOC administered in their agencies and to provide feedback on the process. Three organizations funded by the Local Initiative Funding Partners National Program Office (LIFP) of the Robert Wood Johnson Foundation are included among these organizations. I thank them for taking time from their work to assist in the development of the SOC. A few of these relationships have become long-term and continuing, making their commitment to the process remarkable. Among these, I would like to express my gratitude for the support and encouragement of Larry DeMarzo, executive director of the House of the Crossroads; Steve Galen, executive director of the Primary Care Coalition of Montgomery County; Judy Monahan-Grystar, executive director of Turtle Creek Valley MH/MR; David Fisher, chief executive officer of Oak Orchard Community Health Center; and Connie Mercer, executive director of HomeFront.

I wish to acknowledge the contributions made by Drs. Raymond Logan and Gregory Turner in the development of the SOC. Both spent many hours in developing the rubrics and scoring organizations. They pointed out times when I lost conscious awareness of

my thinking processes and brought me back to the practical necessity of communicating with them. Greg has been especially diligent in pursuing "foundation" articles and introducing relevant literature beyond my ken.

The SABI coordinator, Karen Brodsky, has made a number of site visits with me and spent many hours discussing SOC development issues. Evaluators often talk about partnerships between evaluation and program staff, but it is seldom as comfortable a reality as it has been with Karen. The LIFP deputy directors, Sandra Lopacki and Ted Hardgrove, each traveled to grantee sites to introduce me to grantees and to observe the process of SOC administration. Our discussions have been very helpful to me in refining aspects of the SOC. Site visits with them gave the SOC its first real test and demonstrated that the instrument could document organizational change.

Innovating in the course of doctoral studies can be a frustrating and difficult experience without sound guidance and expert advice. That my dissertation has been a rich and rewarding learning experience is largely attributable to a diverse committee of eminent scholars, each of whom was highly accessible and generous with his or her time and counsel. Dr. Mary Kay Stein, associate professor in the School of Education and senior scientist at the Learning Research and Development Center, brought practical and research experience in cognitive studies to the committee. In addition, she introduced me to Pasteur's Quadrant, in which I aspire to be working. Dr. Louis A. Pingel, associate dean of the School of Education, brought his experience in teaching statistics and research methodology to the committee. Dr. Pingel also taught me the value of drawing diagrams and pictures of the problems I try to solve. Dr. Laura C. Leviton, a nationally

recognized evaluator and my program officer at the Robert Wood Johnson Foundation for the SABI project, brought extensive evaluation research experience, knowledge of evaluation use, and an understanding of the nonprofit philanthropic organizational culture to the committee. She also introduced me to my future dissertation director and to many of the most respected people active in evaluation today through the American Evaluation Association (AEA). Dr. R. Tony Eichelberger, of the Department of Administrative and Policy Studies in the School of Education, brought his practical approach to evaluation design to inform decision-making and his study of disciplined inquiry to the committee. He introduced me to the American Educational Research Association (AERA) and to the AEA. He also introduced me to most of the foundational literature in evaluation, which I bound into three volumes. My relationship with these four professors has evolved from those of student and employee to those of teacher and supervisor and, now, to those of colleagues and friends. Any errors or omissions in this work are mine and mine alone. That there were not more is a tribute to this committee's good work..

Dr. Margaret C. McDonald, my wife and partner, brought more than the usual spousal support to this effort. Her writing and editorial skills gave me the benefit of an additional reader without the concern of what she might ask at the defense. In addition, she added an element of *Integrity* to the realization of this dream.

1. INTRODUCTION

The purpose of this study is to investigate the validity of a maturity model for staging the capacity of nonprofit community organizations to implement or sustain social interventions or demonstration programs. The maturity model is intended to identify and organize the important facets of organizational capacity. The integration of literature from a variety of disciplines and usually disparate fields, from theories of organizations to theories of change, from human resources and role definition to administration and leadership, from cognitive psychology to expertise, and from artificial intelligence to evaluation provides the initial delineation of the various features and sub-features of the model. These were extended by the variety of ways foundation leaders and community agency administrators viewed organizational capacity through the development and initial tryout process.

Researchers in the field have identified a number of instruments to measure organizational capacity. The validity of these instruments, if studied by the authors, is usually determined by field tests showing their perceived value. Some authors provide a rationale for their instrument by identifying selected theory and research they used to develop it. Although informative, this presentation is always general and somewhat misleading about the knowledge on which the instrument is based and the systematic empirical evidence collected. The problem addressed by this study will be to document

the validity of the Staging Organizational Capacity (SOC) maturity model and the specific attributes indicative of developmental stages, based on evidence systematically obtained throughout the development and implementation process. The remainder of this chapter identifies the context from which the organizational capacity of nonprofit community organizations arises as an important part of societal change efforts.

Community organizations have an established role as "laboratories" or "incubators" in the implementation of demonstration programs or the conduct of social experiments to inform policy making or to contribute to understanding of treatment effectiveness (Iscoe & Harris, 1984; Leviton, 1994; Leviton & Schuh, 1999). Funding agencies typically have two major goals for projects in these settings. One is to make meaningful change in the host community or targeted population. The second is to identify or test interventions that developers believe may prove to be effective in other communities to address important social concerns, such as improving health care, reducing health risk, improving HIV education, reducing interpersonal violence, or reducing homelessness. When evaluations conclude that demonstrations or innovations are unsuccessful or have unclear results, decision makers often conclude that the strategy or program is ineffective and redirect their support to other areas. Often, the problem has nothing to do with the innovation or underlying concept being studied but, rather, can be attributed to some difficulty of implementation caused by agency capacity, or by weaknesses in evaluation methods employed.

Small community agencies are frequently unable to implement programs exactly as planned, or to respond rapidly to problems emerging after implementation begins, resulting in a mismatch between the planned and actual intervention. Frequently,

innovations are stated in such broad, general terms that what is being evaluated or how it is being measured is unclear or highly ambiguous. These "molar" constructs are defined as large and complex packages, or interventions, taken as relatively undifferentiated wholes as opposed to molecular or multidimensional constructs that identify the complex array of causal elements that the intervention comprises (Cook & Campbell, 1979; Lipsey, 1988; Shadish, Cook & Campbell, 2002). Such constructs can be problematic for evaluation, resulting in findings that fail to find clear differences between treatment and comparison conditions in social interventions or, if differences are found, providing no insight as to what caused them (Cook & Campbell, 1979). Case management interventions, typically, offer an example of this phenomenon. Within a given case management intervention, the mix of services provided to participants may vary greatly, ranging from locating housing to providing transportation to making appointments. These interventions are frequently found to be effective, but with little explanation of which important components might have been responsible. In the following section, examples of organizational capacity problems in nonprofit community agencies are described. The specific project that required the SOC to be developed is presented.

1.1. Problems of agency capacity.

The Robert Wood Johnson Foundation is one of many funding organizations using nonprofit community agencies extensively as vehicles for conducting demonstrations of innovations in social service delivery in the area of health and health care. Not infrequently, agencies implementing demonstrations fail to implement them as planned or fail to complete them. Several foundations have initiated efforts to attempt to reduce the number of demonstration failures. The Edna McConnell Clark Foundation (EMCF), for

example, has initiated an intervention that makes a substantial investment in agency information system capacity. The Robert Wood Johnson Foundation (RWJF) staff observed that failures in community agencies seemed related to organizational capacity to undertake new or expanded projects. These agencies are often smaller and younger than more established agencies. The value of using such agencies for demonstrations derives, in part, from perceptions that they have closer ties with their constituencies; are service innovators; and are more tightly focused, more quickly responsive, and less expensive than larger, more established agencies working in the same communities (Leviton, 2001; Leviton & Schuh, 1999). These perceptions have not been verified directly through systematic study, but there is indirect evidence that supports some of them. We know, for example, that smaller organizations tend to operate on a more informal basis than older or larger organizations and that informal organization systems tend to be more flexible and more rapidly responsive than formal ones (Downs, 1967; Lawrence & Lorsch, 1969). We also know that community organizations frequently emerge in response to local problems or events, so they can be expected to have very close ties to the community (Light, 2001).

Small agencies sometimes have difficulty in providing the continuity of effort required for successful implementation and completion of demonstrations. Often the problem is one of management's failure to ensure that performance is delivered as expected by the funding bodies. For example, staff turnover is sometimes high in small agencies resulting in delays or interruptions in service. The downside of using community organizations for demonstrations has been studied no more systematically than the

upside. However, nonprofit foundations engaging in demonstration research have been interested in "sustainability" and are now beginning to explore "capacity building" as a partial response to some of these downside issues. The Edna McConnell Clark Foundation, the California Wellness Foundation, the David and Lucile Packard Foundation, and the Robert Wood Johnson Foundation are prominent among those foundations exploring organizational capacity building. Each approaches the problem from a slightly different perspective. For example, the California Wellness Foundation focuses on providing technical assistance (TA) and even defines capacity building in terms of TA. The EMC Foundation is experimenting with investing large sums in a few organizations over the course of several years to determine if increasing capacity will result in better programs or services for children. The Packard foundation has been providing grantees with management consulting and training since 1983, and the Robert Wood Johnson Foundation (RWJF) is experimenting with capacity building prior to grant making or grant expansion in hopes of improving the success rate when community organizations implement demonstration or new programs. The work reported here emanates from the Small Agency Building Initiative (SABI) of the RWJF that began in 1999.

1.2. The Small Agency Building Initiative and its evaluation

The Robert Wood Johnson Foundation invited me to evaluate its project aimed at building small agency capacity. The RWJF staff perspective is that smaller agencies are "...closer to the problems of communities affected by health and social problems than are larger more established agencies," and they tend to lack sufficient "organizational

capacity" to successfully implement agreed upon projects. Therefore, the foundation wished to study the effects of capacity building grants made in addition to, or prior to, demonstration grants, with the hope that such resources might improve the ability of community organizations to implement and complete projects. However, there were three problems: a lack of definition concerning what increased capacity would be, a molar (undifferentiated) intervention construct proposed as the demonstration intervention, and measures of outcome that were weak at best. These problems posed potentially serious obstacles for successful evaluation of the initiative.

1.3. The problem of an under-defined construct, capacity

The proposed intervention was intended to provide a limited amount of technical assistance and infrastructure support to build the capacity of agencies to implement new grants. Capacity was not more specifically defined. Without a clear definition of capacity, it was unclear what the intervention was to accomplish. The potential problem posed for an evaluation is that each organization would receive a different intervention, reducing an evaluation to assessing a collection of unique interventions each with a sample of one. The term “capacity” required an “operational definition” of sufficient precision that it could be used across the grantee agencies with some degree of consistency to assess the extent to which it had been “built” (Blalock, 1972). As the evaluator, I could see that the project would likely produce highly varied results and would be difficult to evaluate successfully due to the lack of an adequate operational definition of capacity.

1.4. The problem of a molar or undifferentiated intervention construct

Molar intervention constructs, like training and technical assistance (T&TA), are likely to be varied from organization to organization. Even if an evaluation detects effects, it will be unable to explain the relative contribution of the various elements of the intervention makes in producing those effects. In other words, it is difficult to determine just what an undifferentiated demonstration intervention has demonstrated. Molar intervention constructs also make replication of the demonstration difficult to achieve (Cook & Campbell, 1979; Shadish, Cook, & Campbell, 2002). The term "organizational capacity" is readily recognized as a molar construct with many interrelated dimensions, as defined by Cook and Campbell (1973). Capacity is a complex, multi-dimensional construct used to describe an array of micro-mediators that may or may not contribute to producing observed effects at a given time. Micro-mediators describe cause and effect relationships on a level smaller than the molar level. In the preceding paragraphs, technical assistance, information systems, and management might all be thought of as different micro-mediators of capacity. Molar constructs can be evaluated, but attribution of observed effects to the interventions studied is highly problematic because similar effects may be produced in a variety of ways. Lipsey (1988) recommends developing more complex constructs to better represent the complexity of the problems being addressed by interventions and to more precisely identify the mediating variables. Clearly, in the absence of pre-existing measures, applying conventional approaches for evaluating the SABI was likely to produce unclear or unconvincing results.

1.5. The problem of anticipating weak outcome measures

Social programs and interventions tend to be weak relative to the problem they purport to address (Rutman, 1980; Yeaton, 1985; Yeaton & Sechrest, 1981) and tend to produce modest results (small to moderate effects) (Tukey et al., 1976; Reichardt, 1994; Sechrest & Yeaton, 1981). At the same time, evaluations are able to detect only moderate to large effects (Crane, 1978; Lipsey et al., 1985; Lipsey, 1990). This combination is a formula for ambiguous or null evaluation results that is frequently overlooked in the planning and evaluation of demonstration projects. Reducing variation in outcomes among sites is one way to sensitize an evaluation to the effects produced by a demonstration. The antithesis also holds. High variability in outcomes introduces noise into the evaluation, making detection of intervention effects more difficult.

The planned interventions (technical assistance) were to be small, of short duration, and varied to meet the unique needs of each agency, suggesting that the expected outcomes themselves would be at best small to moderate and varied by agency.

Reviewing the extensive prior work on organizational capacity revealed no organizing framework adequate for assessing the foundation's initiative. Accordingly, I started to work on developing such a framework and was able to identify a number of dimensions that contribute to organizational capacity to successfully implement new or expanded

innovations or demonstrations. When agency development on one or more of these dimensions is not adequate for the intervention, successful implementation may be problematic.

Since the initial work on the SABI project, I have focused on the development of a maturity model that is called Staging Organizational Capacity (SOC). It has been developed with support from the Robert Wood Johnson Foundation and applied in a number of the Foundation's nonprofit community agency grantees. It has also been applied successfully (i.e., funders and agencies agree with results) in several nonprofit agencies in Pittsburgh; Philadelphia; and Silver Spring, Maryland. This study will describe the processes used to develop the validity of the instrument in a variety of settings. The data obtained and the changes in the SOC that were made to improve its validity and utility in these settings will be used to document that basis for the features, sub-features, and attributes used to describe the stages of development of organizational capacity of nonprofit community agencies. These results will demonstrate the value and applicability of the SOC maturity model. The characteristics of maturity models and the features and sub-features of the SOC follow the problem statement.

1.6. Problem statement

The problem statement that identifies what this study investigates is: What is the validity of the SOC to identify important characteristics of the capacity of small nonprofit organizations?

The instrument was developed as a maturity model (see next section); therefore, assessment of validity addressed the extent to which the maturity model's features, sub-

features, and attributes identify important capacity characteristics. Validity was tested in four ways. (1) The SOC was compared to the Self Assessment Tool (SAT) of organizational capacity, in order to show similarities and differences of the primary features of SOC. (2) The first and last iterations of SOC were compared to provide the bases under which the instrument was developed further. (3) The SOC was tested for interrater reliability in various settings by three members of the development team. (4) Documented feedback from primary stakeholders on the accuracy and utility of the SOC documented and analyzed.

The data obtained from a wide array of development and initial implementation activities were used in the validation study. These include notes from: (1) the initial Small Agency Building Initiative (SABI) experience, (2) iterations of SOC, (3) consultation with experts, and (4) stakeholder feedback. Reliability of three raters indicates that others can be trained to use SOC with nonprofit community agencies. The SOC was developed as a maturity model. A short description of maturity models follows, and the features of organizational capacity used in the SOC are listed.

1.6.1. Maturity models

A maturity model is defined as involving a set of *features* and a related set of *levels* or *stages* (Lesgold, 2003; Lesgold, 2000). A scoring rubric containing attributes of each feature for each stage of development is employed as a guide for assessing the stages of development of the features at the time of observation. Five basic features and several

sub features were identified as appropriate for staging organizational capacity (see Figure 1.1) based on the non-profit and organizational literature.

Features of Organizational Capacity

1. Governance

2. Financial Resources Maturity
   A. Financial Vulnerability

3. Service systems maturity
   A. Vertical Differentiation
   B. Horizontal Differentiation

4. Human Resources Maturity
   A. Service/Management differentiation
   B. Service/Management specialization

5. Internal operations maturity
   A. Management Information Systems
   B. Finance and Budgeting Systems
   C. Human Resource Systems
   D. Communications Systems
   E. Development Systems

Figure 1.1 Maturity Model Features for Staging Organizational Capacity

The maturity model is predicated on the assumption that organizational features will vary, in part, based on each organization's stage of development from less developed capacity to highly developed capacity. Five developmental stages are defined for this model. These are employed for rating organizations from least developed capacity (Stage 1) to most highly developed capacity (Stage 5) for attributes of each feature. Attributes for the features and sub-features in Figure 1.1 are identified and studied in a wide variety

of literatures. The focus of Chapter 2 is to identify the appropriate fields and references in which these have been studied and to determine the extent to which there is convincing evidence supporting developmental pathways or organizational life cycles for use in SOC.

Chapter 2 discusses some more fundamental concepts on which the SOC's development of SOC was based. The underlying concepts of organizations and of interventions on which the SOC was conceptualized are identified. Some of the important organizational concepts are:

1. Organizations as open systems,
2. Organizational life cycles,
3. Organizational capacity,
4. Profit and nonprofit organizational differences.

Some characteristics of interventions that can moderate the potential impact of a demonstration program or social intervention and the ability to detect the effects of programs are:

1. Intervention strength,
2. Dose/response relationships,
3. Intervention maturity

In Chapter 3 the procedures used to investigate the validity of the SOC are delineated. The four primary analyses that were used to focus and structure the study are specified. Important issues are identified and some of the most relevant literature to be analyzed and synthesized is listed. This seemingly linear investigational structure is somewhat misleading, as many of the concepts arise in a wide range of sources, the meaning of which was synthesized for small nonprofits.

Chapter 4 presents the findings of the study and Chapter 5 concludes by discussing lessons learned and the implications for future research.

2. REVIEW OF LITERATURE

Background

This chapter reviews several literatures that identify organizational characteristics, especially those of community agencies, that affect their ability to implement change. These literatures provide the theoretical and empirical basis for the development of the Staging Organizational Capacity maturity model. Although the features and attributes of SOC are primarily derived from observation and experience, the research literature provides a foundation for the conception of organizational capacity as it is delineated in the SOC.

2.1. Organizations and organizational capacity

An attempt to catalog the lessons learned about organizations would exceed the capacity of the author and the scope of this effort. Rather, this section will present that organizational literature upon which the innovations in developing the SOC are predicated. Contributions from systems theory, and in particular open systems theory, are discussed first, followed by an analysis of evidence that organizations mature through cycles or stages. Next, the construct of organizational capacity is examined, and the final topic is differences between profit and nonprofit organizations as they relate to organizational capacity.

2.1.1. Organizations as open systems

Systems theory provides the basic framework from which organizational capacity was approached in developing the SOC. Some systems theory definitions provide a foundation for the SOC logic. A system is defined as "...a set of objects together with relationships between the objects and between their attributes" (Hall & Fagen, 1968). Objects are the component parts of a system, and attributes are the properties of objects and may be infinite in their variety (Hall & Fagen, 1968). This definition is not yet complete because many "systems" meet these broad criteria without being organizations. The solar system and automobile's electrical system are but two from an infinite variety of systems both natural and man-made. Organizations have the further characteristic of being human constructions that are, in many ways, isomorphisms of organic systems. As isomorphisms of natural systems, organizations can be studied, appropriately, by borrowing conceptually from studies of natural systems. Hence, we speak of organizational behavior and adaptability in biological terms. Just as organic systems exist and function in environments, the concept of "organizational environments" has contributed to current understanding of organizational functioning and development. Systems environments are defined by Hall and Fagen (1968) as "...a set of all objects a change in whose attributes affects the system and also those objects whose attributes are changed by the behavior of the system." These key definitions from systems theory form the basic framework for understanding organizations used in this study, which relies

heavily on open systems theory for identification of organizational objects and attributes. Katz and Kahn (1966) identified nine characteristics that are shared by open systems:

1. Importation of energy,

2. Through-put of energy,

3. Output of energy,

4. Cycles of events,

5. Negative entropy,

6. Information input and feedback,

7. Homeostasis,

8. Differentiation,

9. Equifinality.

The first three characteristics are well understood. Organizations survive by *importation of energy*, raw materials or money, from the environment in return for *exporting* some product beneficial to the environment. The *through-put* is the internal process by which the energy imported is transformed into the product. The process of importation, transformation, and exportation is a cyclic process whereby the outputs provide the sources of inputs for the next cycle. The relationship between funding sources and

grantees or contractors comes easily to mind as an example of this cyclic process for those of us in academic settings.

Negative entropy. Entropy is defined as a universal law of nature "... in which all forms of organization move toward disorganization or death" (Katz & Kahn, 1966). However, open systems can overcome the entropic process by importing more energy from the environment than is expended. By storing reserves, organizations can survive crises and forestall disorganization or death. That is, they achieve negative entropy. Katz and Kahn maintain that open social systems, unlike biological systems, can reverse or arrest the entropic process almost indefinitely, becoming indefinitely "sustainable." Accomplishing sustainability requires that organizations *receive information* that signals changes in the input-output cycle in order to make adjustments that preserve a *steady state* of energy exchange between the organization and its environment. Katz and Kahn call this steady state "dynamic homeostasis." The organization is dynamic in the sense that the balance of inputs to outputs is not static, only the ratio of inputs to outputs is steady under homeostasis.

Open systems tend to become more highly *differentiated* as they mature. Greater differentiation is characterized by the replacement of global functions with more specialized functions. We see differentiation in social service organizations as they develop specialized management and service functions as they grow older and larger (Blau, 1970). Katz and Kahn describe a progression during which social organizations

progress from their initial state "... toward the multiplication and elaboration of roles with greater specialization of function." The principle to be extracted from evolutionary differentiation is that open systems, in general, and social organizations, in particular, are initially "... governed by dynamic interaction of their components; later on fixed arrangements and conditions of constraint are established which render the system and its parts more efficient...(von Bertalanffy, 1956)." This progression has led to a body of literature describing organizational life cycles and stages of development that are a linchpin of the SOC framework. The subject of the developmental stages of organizations is more fully examined below.

A final precept from open systems theory that contributes to the logic of our framework is called *equifinality* (Katz & Kahn, 1966). The principle of equifinality states that "...a system can reach the same final state from differing initial conditions and by a variety of paths" (Katz & Kahn, 1966). A corollary emerging from the principles of differentiation and equifinality is that as organizations become more highly differentiated, the amount of equifinality decreases. Smaller, less differentiated organizations can, therefore, be expected to display greater variety in their internal and external arrangements than larger, more highly differentiated organizations. The implication for evaluation of the Small Agency Building Initiative and development of SOC is that the capacity of larger, more established organizations may be easier to compare and contrast than that of less differentiated organizations.

2.1.2. Organizational life-cycles

Evidence that organizations start, develop, and end along general pathways has been accumulating since just before World War I (von Bertalanffy, 1968; Katz & Kahn, 1966; Downs, 1967; Quinn & Cameron, 1983). Early concepts emerged in neoclassical organizational theory and in the structural/functional work of the modern period (Shafritz & Ott, 1987). The body of this early work focuses on larger and more mature organizations and on their internal workings. The perspective tended to be cross-sectional and static. This limited perspective has evolved to include studies of the dynamics of organizational growth and development, leading to the emergence of systems theory. Applying the precepts of systems theory to the study of organizations has resulted in postulates of some general and predictable patterns of organizational maturity. Increasingly, authors are describing characteristics of organizational life cycles and stages of organizational maturation (Quinn & Cameron, 1983; Downs, 1967; Katz & Kahn, 1966; Light, 1998; Stevens, 1992; Kazanjian & Drazin, 1990; Smith et al., 1985; Prochaska et al., 2001; Crosby, 1980). Most postulate a temporal sequence of development employing from three to seven stages. Katz and Kahn (1966), working from a systems perspective, described the three stages mentioned above. Anthony Downs (1967), in his study of bureaucracy, also used a three-stage model to describe "the life cycle of bureaus." More recently Quinn and Cameron (1983) studied nine organizational life-cycle models representing diverse perspectives. They concluded that there was sufficient evidence of underlying patterns to support a synthesis of the life-cycle models. The patterns they cite include: evidence of a sequential developmental

process, evidence that organizational maturation is a hierarchical progression, and evidence that the same patterns emerge in a wide range of settings, structures, and functions. Their synthesis produced a four-stage model :

Stage 1. "entrepreneurial" stage progressing to a
Stage 2. "collectivity" stage, then to a
Stage 3. "formalization and control stage," followed by an
Stage 4. "elaboration of structure" stage

Kazanjian (1988) proposed a "stage-of-growth model" for technology-based firms, especially new ventures. The four stages of this model, which resembles Quinn and Cameron's,. are:

Stage 1. Conception and Development
Stage 2. Commercialization
Stage 3. Growth
Stage 4. Stability

Kazanjian's model was employed to investigate the degree to which the problems faced by firms studied were associated with their stage of growth. He found "partial" support for a model of predicable patterns of growth. Use of the model was extended to study, in 105 companies, the relationships by stage of development for decision-making, functional specialization, and rate of growth in 105 companies (Kazanjian & Drazin, 1990). This study is important, not only for reporting positive findings, but for being one of the few empirical studies of stage-state relationships. A limitation of the study

reported by the authors is noteworthy for the current project, namely that ther model is limited to a cross-sectional analysis and did not support a longitudinal analysis as well.

The empirical work in the domain of organizational maturity is largely descriptive and exploratory in nature. The topics addressed range widely, and models of developmental stages vary from study to study. However, most of the stage-growth related organizational literature does not describe empirical research but, rather, is prescriptive or explanatory in nature. Fletcher and Taplin (1999) employ a six-stage maturity model to provide leadership guidance to managers in a "food-for-thought" genre book in which management styles and techniques are asserted to be related to an organization's evolutionary stage. Crosby (1980) employs a five-stage maturity model for measuring "quality management." This is a "how-to-do-it-yourself" genre work for assessing the stage of development of "quality management." No claim is made that the resulting measures have specific meaning. However, the authors do assert that the instrument can be used to compare companies or divisions. Simon and Donovan (2001) employ a five stage maturity model for use by nonprofits in assessing the stage of development of their organizations. This is also a "do-it-yourself" genre work that, in my experience, appears to be somewhat better grounded than many works of its kind. Five stages are assessed along seven organizational features like governance and leadership. A scoring rubric is provided for those using the instrument to rate 15 items for each feature on a five point Likert scale ranging from "least like us" to "most like us." This work shares more in appearance with the SOC approach than any other model or instrument reviewed. The similarities and differences will be discussed later in the section on evaluation.

2.1.3. Organization capacity and capacity-building as molar constructs

Capacity-building is a term that has become commonplace in the evaluation world. Funding bodies want to implement and evaluate capacity-building in community, organizational, family, and individual settings (Scheirer & Thayer, 2000; Wilder Foundation, 2000; the California Wellness Foundation, 2001; Letts et al., 1999). The term capacity-building is variably defined as illustrated by the following citations from the literature:

- Capacity building is "...enhancing the ability of individuals and groups to mobilize and develop resources, skills and committments needed to accomplish shared goals." (The Wilder Foundation, 2000).

- "Capacity building is the development of an organization's core skills and capabilities, such as leadership, management, finance, and fundraising, programs and evaluation, in order to build the organization's effectiveness and sustainability." (California Wellness Foundation, 2001).

- "And by 'capacity-building' we mean building the capacity to fulfill an organization's mission." (Jacobs, 2001).

 “Capacity building can be defined as structural changes in an organization like the addition or re-allocation of staff, the alteration or creation of an MIS, or the use of consultants” (Vinzant & Vinzant, 1996).

 “Building [evaluation] capacity is thus the enterprise of raising the_collective_ability to perform a certain level of evaluation tasks ...." (Stevenson et al., 2000).

Organizational capacity is not usually considered in the abstract. Letts et al. (1999) talk about program delivery capacity, program expansion capacity, and adaptive capacity. Harrison (1987) discusses organizational capacity to implement change. Fredericksen and London (2000) discuss organizational capacity to administer projects. Stevenson et al. (2000) discuss the evaluation capacity of organizations. As can be seen, the term capacity in reference to organization capacity or capacity building is a molecular construct comprising elements that vary both in kind and number depending on use.

Many foundations use the term capacity or capacity-building to describe either "sustainability" of an organization or an intervention for the term of a grant, or for the period after which grant funds are no longer available. The term capacity building is used frequently to describe efforts intended to improve sustainability of either organizations or specific demonstrations. Strategic planning, fund raising, or communications are believed to be some of the mechanisms to achieve this improvement,

and capacity building interventions tend to target increasing or improving staff skills in these domains. Technical assistance (TA) of some form is the vehicle for delivering the interventions, usually on staff skills (California Wellness Foundaton, 2001; The Robert Wood Johnson Foundation, 2000). TA has been defined as "... an individualized educational program." (Reifler, 2000).

My experience is that limiting capacity building interventions to TA obscures vital dimensions of the construct of capacity like the collective abilities noted by Scriven and the structural resources noted by Vinzant and Vinzant {1996). Therefore, I construe capacity as comprised of both *organizational resources* and *individual skills*. Individual skills and experience consist of the human capital assets of an agency and are an indicator of how well an agency can perform. Organizational resources include the structures, systems, and resources of an agency. Organizational resources influence how individuals within an agency can work. Agency productivity is a function of the skills and experience of employees *plus* those resources that organizations bring to bear permitting individuals to function at various levels of skill and expertise (Letts, Ryan, & Grossman, 1999). Highly productive agencies will have organizational resources and arrangements sufficient to permit employees and volunteers to be assigned to tasks requiring their greatest skills. Less productive agencies may have people working on tasks that do not require their best skills or they may have people assigned to tasks for which their skills are less than ideally suited.

2.1.4. Differences between profit and nonprofit organizations

A debate about the similarities and differences between managing and operating profit making organizations and nonprofit service organizations has been taking place for many years (Zaltman, 1979; Letts, et al., 1999). Currently in vogue are a host of private sector management practices touted as improving nonprofit performance. Many consultants and policy makers assert that applying private sector practices that to contribute to success in business would improve nonprofit organization productivity and sustainability. I have been unable to find research that supports such assertions. However, there is evidence suggesting that such mimetic isomorphism is as likely to introduce dysfunction as it is function (DiMaggio & Powell, 1983). Indeed, a variety of profit-world practices have been attempted in the nonprofit environment with highly mixed results and most have only limited periods of popularity. Some of the practices tried in nonprofit settings and found to be flawed or of limited value include Total Quality Management (TQM) (Senge, 1990), strategic planning (Light, 1998; Hamel & Prahalad, 1989; Mintzberg, 1994), and an array of performance measurement practices (Perrin, 1998). Repeated failure to replicate profit world-results in nonprofit environments suggests the existence of some important differences between the two types of organization that might be fundamental. Three of these differences contribute to the underlying logic of this effort.

First, nonprofits differ from profit sector enterprises in that their end is not to generate profits or monetary surpluses. Profit-making organizations develop a product or service first and then bring it to market. The infrastructure necessary to raise capital and develop

the organization's product and bring it to market must exist first. Non-profit organizations, on the other hand, frequently come into being to provide services for which there is already a market. The consequence is that nonprofits, especially smaller ones, are apt to lack the infrastructure and resources to optimally manage their projects. In fact, non-profits are characterized as being part of a culture that eschews developing the infrastructure necessary to become highly productive (Letts, 1999). Policy makers, funding sources, and even service recipients tend to view infrastructure development as diverting limited resources from services. In my experience, nonprofit organizations are frequently told, by policy or regulation, how much they may charge for the services they provide, with little or no consideration for the actual costs incurred. They are frequently also told how their income must be spent. For example, foundations frequently proclaim a maximum allowance for overhead or indirect expenditures. The Robert Wood Johnson Foundation (2001), for example, limited overhead to nine per cent of direct costs without consideration of agency size or age. While this practice is consistent with the nonprofit culture described by Letts and company, it may constitute an unintended penalty for smaller organizations. In working with community-based organizations, I have observed that well differentiated smaller agencies tend to have higher overhead requirements as a proportion of total revenues than do larger organizations, and that recovery of overhead costs is a general problem of smaller organizations (Schuh, Leviton, & Stagg, 2000). The problem of inadequate recovery of overhead can be exacerbated when managerial requirements are imposed without adequate consideration of the increased burden. For example a Philadelphia-based foundation funded a multi-organization initiative at modest levels to implement activities intended to reduce or prevent violence. After agencies

were funded, a new responsibility to participate in an initiative-wide management information reporting system was mandated. However, additional resources to support participation were not provided, and agencies consistently reported to this evaluator that the additional burden detracted considerably from their ability to implement activities as originally planned.

The requirement or expectation that agencies develop a strategic plan without consideration of the infrastructure requirements necessary to support such planning is another example of a common mandate on small organizations without provision for meeting the overhead expenses incurred. Vinzant and Vinzant (1996) assert that successful implementation of strategic management practices is related to the resources and the degree of autonomy organizations possess in implementing change. Organizations without adequate infrastructure are not likely to benefit from strategic management. Limited organizational autonomy constrains implementation of strategic management processes. And as the number of constraints increases, the usefulness of strategic management diminishes. The view of these authors is that strategic management practices are less appropriate for smaller organizations supported by grants or contracts to provide services over which they have only limited power to change or alter. Most of the small agencies that I have visited in the past three years indicate that they have a strategic plan, and most report that the exercise of developing the plan was a good exercise. But they also report that the plan has not been followed, is out-dated, or has been shelved in order to attend to issues not covered in their plan. Yet strategic

planning is a basic technical assistance offering to small agencies with little assessment of their capacity to implement and reap the benefits of strategic management.

Second, profit making organizations develop a product and enter the market place to compete for returns on their investment. Nonprofits, on the other hand, are frequently required to respond to a policy initiative or meet the needs of a target population without a prior investment in product (intervention) development. Instead, development takes place concurrently with implementation of the intervention. The mechanism of going to market with an undeveloped or incomplete intervention is part of the problem addressed by Scriven's (1967) distinction between formative and summative evaluation, which was predicated on a perception that interventions are frequently evaluated prior to being developed for optimal performance. The concept of evaluability assessment (Horst et al., 1974; Wholley, 1976,1987,1994; Rutman, 1980) supported and extended Scriven's observation to include social program implementation in general. Evaluability assessment recognizes that programs need to mature to a point of intervention stability before they are able to support an evaluation. The ongoing need for formative evaluation and for evaluability assessments, however, is an indication that, unlike profit making organizations, nonprofit organizations seldom have the luxury of developing an optimal product before it is brought to market and judged by those funding it.

Third, performance measurement and strategic management are of more *practical value* when used with profit-making organizations than with nonprofits. A reason for this

difference relates to the nature of the products. Profit making ventures can relate their income directly to their output and assess the degree to which they meet expectations. For example, McDonald's or Wendy's can count the products they sell and the return derived. Nonprofits, on the other hand, are engaged in providing services. Their output is essentially measured in terms of the activities completed in providing services, frequently on a fee-for-service basis. Activity data are inherently flawed. Activity data tend to inflate over time (Blau, 1963) and to become distorted as they are reported (Downs, 1967), making activity measures poor candidates for assessing agency effectiveness, in general (Perrin, 1998).

Summary. Organizations can be characterized meaningfully as open systems existing in an environment from which they must exchange energy in order to survive. Organizations can be viewed as maturing by passing through developmental stages of growth that are generally sequential in nature. Organizations exhibit similar structural and functional characteristics at a given stage of development. There is no general agreement on the number of stages appropriate for studying organizational maturation. However, three, four, and five-stage models are most prevalent.

The concept of organizational capacity is a molar construct that can comprise a highly variable number and type of elements. Capacity includes both individual and organizational attributes. Individual attributes include the combined potentials (skills, knowledge, and experiences) that those working in an organization bring to their jobs.

Organizational attributes include the resources (structures and technologies) that condition how individuals are able to apply their skills and experiences to their work.

Finally, this section presented some evidence of differences between profit and nonprofit organizations that might make the application of management practices considered effective in the profit world irrelevant when applied to nonprofit organizations. Nonprofits tend to have underdeveloped or sparse management infrastructures compared to profit-making ventures. The relative lack of structural arrangements has the effect of limiting the comparative adaptability of community organizations. Social service organizations do not have the same autonomy as profit-making organizations. They are more directly regulated in what they may charge for their services and even how they deliver these services. It has been shown that, as autonomy decreases, the value of strategic management decreases (Vinzant & Vinzant, 1996). In general then, strategic planning may be less appropriate for the nonprofit world than for the profit-world. Typically, the productivity of social service organizations is measured in terms of activities. We know that activity data have the inherent flaw of inflation and distortion through time. This flaw limits the utility of using profit world management techniques like TQM and performance measurement in much of the nonprofit sector.

In this section we have considered lessons from the literature on organizations because they are the targets of capacity-building interventions intended to change or improve them. Successful evaluation of the effects of such interventions requires an

understanding of the nature of intervention factors that can influence success or failure to either produce or detect the intended change. The next section turns to a consideration of assumptions about interventions that have contributed to the rationale for developing the SOC.

2.2. Interventions

In this section aspects of social interventions are discussed from three perspectives; (a) social interventions as delivered by social service, educational, and treatment organizations, (b) the role of demonstrations as subset of (a), and (c) interventions intended to produce changes in organizations or organizational capacity.

2.2.1. Social intervention

Social intervention, as used here, is synonymous with social programs as defined by Weiss (1972), includes treatments as intended by Lipsey (1990) in discussing treatment effectiveness research, and also includes programs, projects, and elements as defined by Cook, Leviton, and Shadish (1985). Two attributes of interventions will be examined: the tendency for interventions to be weak, and evidence that they pass through developmental stages.

1. *Intervention strength*. In general, social interventions are not of sufficient strength to accomplish their intended purpose. The clearest statement of the fundamental problem was made by Charles Reichardt (1994) when he wrote:

For the most part, our social programs are not solving the problems to which they are addressed. One of the major reasons for these failures is that social programs tend to be

puny and poorly developed. By puny, I mean the resources devoted to social programs are small relative to the size of the problem. By poorly developed, I mean relatively little effort goes into the program's conceptual planning and design.

This provocative statement reinforces a similarly provocative observation by Rutman (1980) a decade earlier that:

The causal assumptions linking a program to some of its goals are often weak. In these instances the program represents a puny response to complex and long-standing problems.

Intervention weakness or strength is defined as a function of the relationship between the amount of resources devoted to addressing a problem and the size of the problem itself. Alternatively, it can be interpreted as a comparison of the complexity of the problem with the complexity of the intervention addressing it.

Comparing intervention resources to the magnitude of the problem can become a multi-level undertaking. On an individual level, the resources available may or may not be sufficient for those being served. For example, some HMOs place a limit on the number of psychiatric visits to which patients are entitled. That limit is frequently on the order of 28 to 30 visits either per year or per lifetime depending on the HMO. On the average,

this arbitrary number might be adequate, meaning for half of those in need it is likely to be inadequate. Interventions to treat anxiety or depression might be more successful within a 30-visit limitation than will those aimed at treating the homeless who have mental health or substance use problems.

A second example at the organization level is the Partner's Project, designed to treat mothers in recovery from substance abuse and their prenatally exposed children. A major component of that project was supervised sheltered housing. The project always had a waiting list of potential clients because of the limited number of apartments available to the program. This project, successful on an individual level, was inadequate to address the magnitude of the problem in its community, with a waiting list months long.

References addressing intervention strength are slowly emerging in the evaluation literature under the leadership of Lee Sechrest (1979). Employing a medical analogy, Sechrest introduced the concept of intervention dosage, noting that if the dose was not strong enough for the seriousness of the problem, the treatment was not likely to be successful. Yeaton and Sechrest (1981) observed that treatment strength is a function of both the planned intervention and the integrity with which it is administered. Sechrest and Yeaton define treatment strength as "...the a priori likelihood that a treatment would have it intended effect." (Yeaton, 1985). Variation in intervention strength can be related to differences in intensity (Reicken, 1976), duration (Weiss, 1972), or rate of decay (Lipsey, 1990). As of this writing, no more meaningful synthesis of intervention strength

has emerged. However, interventions of a given strength are not likely to have a uniform effect on individuals, organizations, or communities. Readiness or capacity of a target to benefit from a given intervention can be expected to vary in two important ways. First, individual differences in response to an given intervention can be expected. Termed the *dose/response* relationship, this is a public health/medical concept that is increasingly being applied to social interventions (Yeaton & Sechrest, 1981; Lipsey, 1990; Hermann, 1998). Second, *stage appropriate* interventions have been found to be more successful than more general and stage inappropriate interventions (Prochaska & DiClemente, 1992).

2. Dose\Response relationships. A positive relationship has been reported between the amount of treatment and the amount of resulting benefit (Howard, et al., 1986; Yeaton & Sechrest, 1981; Lipsey, et al., 1985; Test, 1981). Generally speaking, stronger interventions are more successful and weaker interventions are less successful. Sechrest et al. (1981) cite three potential reasons for weak treatment effects: Weak interventions (treatments), interventions inappropriate for conditions, and weak dosage of a potentially strong intervention. Prescriptions for improving intervention results generally focus on strengthening the dose as a means for strengthening results.

When *dosage* is not appropriate for the seriousness of the problem in a given intervention, the treatment, predictably, is not likely to induce the intended response. Increasing the treatment might produce a better response if the increase is tailored to the individual response. There is evidence that more flexible approaches might improve the

success of interventions. For example, reports from community support programs have established that comprehensive programs to provide community care for the mentally ill are more successful than less comprehensive ones. The Community Support Program (CSP) has been one of the most extensively researched social programs, and findings supporting the efficacy of comprehensive approaches have been replicated. However, the reasons for the result have not been reported or even fully understood. My experience in evaluating case management systems leads me to believe that the reason may be related to comprehensive programs possessing the flexibility to tailor intervention strength for each client. We had a similar experience with the Integrated Learning Project in which we provided enhanced support for Welfare to Work (WtW) mothers identified as having mental health problems. By providing short-term support and additional tutoring, the drop out rate from training was reduced by nearly 50%.

The dose part of the dose/response relationship has been the focus to this point. Dosage is a function of the intervention and those administering it. Now let us turn to a brief consideration of the response part of the relationship.

The intended targets of an intervention may vary in their *response* to interventions of similar strength. Individuals or organizations vary in their readiness to change. Those more amenable to the intervention may require less to produce the desired outcome while those less amenable may require more (Schuh & Leviton,1999). However, most government or foundation sponsored social programs tend to come in a one-size-fits-all

package, with the object of serving greater numbers at the expense of accounting for individual differences (Yeaton & Sechrest, 1981). The obvious pitfall to the normal approach is that only those whose profile approaches the median are likely to be best served, while any on the extremes are likely to be less well served (Schuh & Leviton, 1999). Individuals requiring more or stronger intervention than that delivered can be expected to be little changed while resources may well be wasted on those requiring less than delivered to produce the intended effect.

Social interventions are intended to produce some change or improvement in people, institutions, or communities (Weiss, 1972). Intervention design and delivery, as discussed above, relates to the agents of change. Adequate intervention design and delivery are necessary but insufficient conditions for successful change. Readiness for change on the part of the target of change also conditions the likelihood that interventions will succeed or fail. The best interventions are unlikely to produce the expected results in noncompliant recipients. Targets of interventions can be expected to vary in their readiness to change. Evidence that change and readiness for change occur in stages has been accumulating for the past decade. Prochaska, DiClemente, and Norcross (1992) developed their "transtheoretical model" to assess readiness of clients to overcome addiction. They developed Stages of Change assessment and, after a decade of use and associated research, concluded that experience supported their readiness for change stages and that stage-appropriate interventions are more effective than are more general addiction treatment interventions. Evidence supporting effectiveness of stage-tailored interventions exists in other areas as well. Welfare-to-Work interventions report better training program completion rates through use of profiling techniques to identify

participant stage of readiness to work and targeting intervention resources appropriately (Ebert, 1997). Organizational structures, activities, and criteria of effectiveness vary by stage of maturation with the implication that interventions appropriate for one stage might not be appropriate for other stages (Quinn & Cameron, 1983). Vinzant and Vinzant (1996) assert that the chances of successfully implementing strategic planning are improved by ensuring that the management approach is appropriate for the developmental stage of organizational capacity. A study by Kazanjian and Drazin (1990) gave support to the superiority of stage-contingent variables over organizational variables in studying issues of organizational structure and process. This evidence suggests that stage-appropriate interventions tend to be stronger and more successful than are more general, or one-size-fits-all, intervention designs.

The dose/response social intervention literature emerged from clinical interventions and has been highly concentrated on individual level responses (Sechrest et al., 1981). However, the literature does appear to justify expanding the concept to include interventions targeting organizations and communities. In fact, DiClemente and Prochaska's stages of change model for addiction is now being adapted for use with organizational change interventions (Prochaska, Prochaska, & Levesque, 2001).

It is widely acknowledged that social interventions tend to produce only modest results or effects (Rutman, 1980; Tukey, 1976; Reichardt, 1994). And it is evident that stronger doses of intervention have the potential for producing greater responses. Therefore, it is

perplexing that interventions do not discuss dose/response relationships more frequently and directly in their planning and evaluation reports.

3. Intervention maturity. Interventions have a developmental aspect, and their stage of development has implications for their effectiveness. Scriven alluded to this developmental nature when he distinguished formative from summative evaluation. Formative evaluation is used to optimize or perfect the evaluand while summative evaluation is used to judge effectiveness (Scriven, 1967). In making this distinction, Scriven recognizes that interventions are not born fully mature and are likely to have their greatest effects only after a period of development. Scriven's emphasis was on product development (Shadish, Cook, & Leviton, 1991) and so formative evaluation addressed, in part, issues of intervention design. The mechanisms for delivering interventions can also influence intervention success (Scheirer, 1987). Interventions sometimes need start-up time in order make staffing arrangements or infrastructure adjustments before becoming fully implemented; and until full implementation is achieved, an evaluation is not likely to find the intended or planned results. Clearly, better developed and implemented interventions are more likely to produce the planned results than less mature interventions.

2.2.2. Demonstration programs.

Demonstrations or demonstration programs are a special category of social intervention concerned with new ideas or innovations that might improve existing practice. Two

important attributes of demonstrations have influenced the development of the SOC: first, social demonstrations are intrinsically linked to evaluation research; and second, successful demonstrations have a temporal order of development and that order has implications both for intervention implementation and for evaluation research methods.

That demonstrations contain an exploratory or knowledge-seeking element can be inferred from the definition of demonstration. Demonstrations are undertaken to show that something works or is practical. Some demonstrations are undertaken to compare innovations to accepted or current practice. Without the aspect of showing or learning how something works a demonstration is not a demonstration. As Suchman (1970) put it, "a demonstration without evaluation is not a demonstration."

Stages of demonstration program development were first described by Edward Suchman (1970), who differentiated demonstration programs by varying purpose into pilots, models, and prototypes to which Weiss (1972) added an institutional stage to accommodate demonstration programs that become operational programs. The literature presents evidence to support inclusion of a fifth stage, an exploratory stage. Therefore, a five stage model of demonstration program or project maturity is described here.

Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
Exploratory	Pilot	Model	Prototype	Institutional

Figure 2.1 Demonstration Research Maturation Model

The exploratory stage of a demonstration has been described most directly by Tukey. On the importance of exploratory work to social science, he observed in his book on exploratory data analysis (1977), "It is important to understand what you CAN DO before you learn to measure how WELL you seem to have DONE it."

At a conference convened to examine the state of evaluation research, Tukey prefaced a presentation on social experiments by pointing out that social experiments represented studies confirming pre-experimental exploratory and pilot work (1976). He observed that the pre-experimental work necessary to support a successful social experiment takes longer and is more difficult and more expensive to complete than are studies confirming the findings of such work. This exploratory work emerges from practice and represents the insights, observations, and hunches of those working in a field about innovations that might improve results. The objective at this stage is to develop innovations in a form that can be piloted. The Integrated Learning Project (ILP) provides an example of an exploratory stage. The project was intended to provide supplemental adult literacy training to WtW participants enrolled in vocational training. We knew that many of our client population were likely to have mental health or substance abuse issues that could influence the expected outcomes. We, therefore, established an informal process of identifying those who might have such problems and providing them with additional intervention support with the result of improving the trainee completion rate (Schuh et al., 1999). This exploratory work led us to seek a more formal pilot for identifying WtW

participants with mental health and substance abuse issues for whom more intensive short term intervention would be appropriate.

The pilot stage is described by Suchman (1970) as a trial and error stage, during which innovations can be tried out and perfected. At this stage issues like the dose\response relationships of interventions are explored. This is also the stage at which the particulars of implementation should be reconciled with the intent of the innovation. The object of a pilot stage is to develop a stable model intervention that can be tested. The ILP example above provides an illustration of a pilot emerging from exploratory work. A more formal assessment and short-term intervention was piloted with a second cohort of WtW trainees. For the pilot, a senior clinician with extensive experience in assessing and treating clients with mental health and substance use was employed to provide more accurate identification of potential problems and more appropriate intervention or referral for services. Lessons from the pilot were employed in developing a model for testing. For example, we learned that assessment and intervention worked best when participants did not need to leave the training site for them. We learned that many WtW mothers could benefit from an intervention aimed at helping them make a transition to a new set of expectations and discipline necessary to complete work training. These lessons were combined to develop a model for testing.

The model stage emerges from a pilot or series of pilots and has as its object validating an approach or confirming insights gained from pilots. This is the confirmatory process described by Tukey (1976). Work at this stage seeks to determine whether the model can

be generalized beyond the pilot experience. Using more controlled conditions than possible or appropriate for pilot work, model testing determines the extent to which pilot results are confirmed. This is the stage at which the rigor of experimental methods is most important and most appropriate (Suchman, 1970; Weiss, 1972, 2001). In fact, Weiss (1972) asserts that this is the only stage of demonstration maturity at which experimental methods are necessary, being premature and, therefore, inappropriate for the prior stages and unnecessary for the following stages. This assertion will be revisited in greater detail in the discussion of evaluation issues in the next section.

Successful model tests require interventions that are stable and implemented as designed. Failure to achieve intervention stability indicates insufficient formative work necessary to support an experiment during the exploratory and pilot stages (Schuh, 1998)

If the model intervention is successful under controlled conditions, the next step is to implement *prototypes* of the model under less controlled conditions to see whether it produces the intended results under real-world conditions. During this prototype stage implementation may take place at several locations or under more varied conditions. The object is to determine the practicality of broader implementation or to explore conditions necessary for success in the broader context.

Successful prototypes can be expected to be developed into operational programs or *"institutionalized*" as a matter of policy. The institutional stage can be reached by

interventions becoming identifiable programs or through their integration as components or options to existing operational programs.

Suchman's model of stages of development for demonstration projects bears a striking resemblance to the developmental stages of a medical clinical trial lending it a degree of credibility (Evans & Ilstad, 2001). That similarity has not gone unnoticed as social interventions with medical patients have been evaluated and reported as "clinical trials" (Wallance & Liberman, 1985). A clinical trial analogy has aslo been employed for describing evaluation of other social programs (Passell, 1993) and it appears to be a useful construct for understanding many of the issues relating to successful demonstration projects or programs.

2.3. Evaluation

Aspects of organizations that might influence organizational development interventions were examined in Section 2.1. Organizations, in this case, are the target of innovation. Issues relating to social interventions (treatments) as they influence outcomes were examined above in Section 2.2. This section presents issues pertaining to the role of evaluation (analogous to diagnosis or examination) as they relate to the determination of success of demonstrations or innovations. Evaluation concerns that provide a foundation for the need to develop the SOC include:

1. Many evaluators believe that their craft is perceived as ineffective.

2. Evaluation methods and measures employed can mean as much to the perceived success of a demonstration or innovation as the innovation itself.
3. Evaluation of projects in the formative stage must be highly sensitive to subtle change.

A discussion of these issues follows.

2.3.1. The failure of evaluation to document success

It is widely perceived that evaluation or applied social science is relatively ineffective in assessing intervention outcomes (Donaldson, 2001; Lipsey, 1985; Boruch & Wortman, 1979; Campbell & Stanley, 1963). As a result, many evaluators conclude that there is little evidence that social interventions produced the intended results. Social innovations tend to produce small or moderate effects for many of the reasons discussed earlier. Evaluations of social innovations, on the other hand, tend to detect effects only when they are moderate to large. The mismatch between the size of the effects typically produced by interventions and the sensitivity of our tools for measuring or detecting those effects contributes considerably to the failure of many evaluations to demonstrate program success (Lipsey, 1990; Reichardt, 1994; Tukey, 1976).

2.3.2. Contribution of evaluation to successful assessment of innovations

Clearly, since the case is that demonstrations or innovations can be expected to produce small to moderate results and evaluation can only detect moderate to large effects, the

prospect looms large that evaluation will fail to detect many positive intervention effects. The prospect of not providing a fair test of social interventions ought to be of grave concern to those who implement or evaluate them. All too often, I have heard program staff, when confronted with a null evaluation, say that they "could not prove it, but they knew that they were having positive results." Although such may not always be the case, the question becomes, "What is the likelihood that well intended evaluations will fail to detect positive intervention effects?" The answer can be inferred from some work of Mark Lipsey (1997), who studied the role of methods in outcome evaluation. He found that the evaluation factors (methods and sampling error) were twice as likely to contribute to variations in observed effects as were the interventions themselves. Evaluation methods alone accounted for about the same variation in effects as the interventions themselves. His studies were based on meta-analyses of hundreds of evaluations, but true to his own advice to evaluators, he employed multiple lines of evidence (Lipsey, 1981) and found similar patterns for studies of more than one intervention area. Lipsey found a third source of variation in effects attributable to issues of measurement contributing nearly as much variation as either the intervention or evaluation methods. One might reasonably conclude that evaluation issues contribute more to the findings of success or failure of many innovations as the innovations themselves. The conservative lesson to be derived is that improving evaluation results requires at least equal attention to *evaluation* design, methods, and implementation as to *intervention* design, methods, and implementation and the characteristics of the implementing agency.

2.3.3. Evaluation of formative projects must be sensitive to subtle changes

Since social interventions can be expected to produce modest to small effects and since evaluations can be expected to detect only modest to large effects, the burden increases with formative projects. Scriven (1967) coined the term formative evaluation to describe situations in which the projects or interventions were still undergoing development. He pointed out that such interventions were not producing optimally, and so the program of intervention effects might be even smaller than those planned. One burden for evaluation of formative projects is simply to be able to detect that a change has been produced regardless of the statistical significance (Lipsey, et al. 1985). Recognition of this evaluation burden has been recurring in the literature for much of the last 50 years (Campbell & Stanley, 1963; Suchman, 1967, Scriven,1967; Sechrest & Yeaton, 1981;Lipsey, et al., 1985). Among the recommendations for increasing the sensitivity of of formative evaluation are (1) increase the power of evaluation to detect effects (Crane, 1978; Lipsey, 1990), (2) increase exploratory work before designing evaluation (Rutman, 1980; Tukey, et al.,1979;. Leinhardt & Leinhardt, 1980), and (3) use of observational techniques (Cooley, 1978; Concato & Horwith, 2000).

In the next chapter the procedures used to investigate the theoretical and empirical bases for the features, sub-features, and attributes fo SOC are delineated. How the maturity model arose from previous stage models and organizational life cycles will also be investigated.

3. METHOD

The purpose of this study is to investigate the validity of a maturity model for staging the organizational capacity of nonprofit community organizations to implement or sustain social interventions or demonstration programs. The problem of this study is to investigate the validity of the SOC maturity model and the specific attributes indicative of stages of development of nonprofit community agencies. Evidence supporting the validity of SOC has been accumulating from several sources, including the exploratory development process of SOC; feedback from evaluators, agency staff, and RWJF staff; and application of SOC in the field. The study of this evidence is organized around four analyses: (a) content validation by comparison of an early version of new instrument with existing instruments assessing capacity to determine the extent to which it assesses important dimensions of organizational capacity as conventially viewed, (b) construct validation by comparison of an early version of the model with a later version to assess its evolution, (c) estimates of interrater reliability among three raters, and (d) construct validation through feedback from agency staff and governing board members and from staff involved in funding those agencies.

One of several techniques is frequently used for investigating the validity of new instruments. Four kinds of validity studies have persisted for the past half-century; content validity, criterion-related validity, construct validity, and face validity (Mosier, 1947; Cronbach & Meehl, 1955; Huck, 2004). None of these was found to be ideally suited or entirely appropriate for estimating the validity of the SOC owing to its unique

nature as a maturity model using observational methods and still in formative stages of development. With no single ideal method available, a mixed methods approach was employed to determine the extent to which the results converged to establish the validity of the SOC in accordance with Lipsey and colleagues that when any one line of evidence is likely to be flawed, a convergence of several lines of evidence are more convincing than any single one (Lipsey, et al., 1981). The study of SOC validity will comprise evidence from four sources: (a) the initial development of the central features of organizational capacity and the maturity model (content validity), (b) an analysis of the revisions of the initial SOC based on field trials in SABI grantees and in nonprofit community agencies in Pittsburgh (construct validity), and (c) initial assessment of interrater reliability among the three SOC project staff (reliability), and (d) feedback from RWJF staff and agency officials (construct and face validity). These are described in greater detail below.

3.1. Data sources

Development of the SOC has been an ongoing process for the past three years, during which extensive records on the process have been maintained. The data fall into four categories; (1) notes from the Small Agency Building Initiative (SABI) startup, (2) various iterations of the SOC during its development, (3) consultation with evaluators having expertise in community and organizational capacity building, and (4) feedback from agency and foundation stakeholders as the SOC has been applied. The relevance of, and contribution of, these sources of data to the validation of the SOC are described in greater detail below.

1. *Small Agency Building Initiative notes.* As indicated in Chapter 1, the SABI project

was an initiative of the Robert Wood Johnson Foundation concerned with building the capacity of small nonprofit community agencies that were grantees or potential grantees. The project began in 1999, and I was involved from the outset in the role of evaluation consultant and, later, as evaluator. During the first year of the project, SABI and RWJF were in frequent contact by phone, email, and meetings as we worked to operationalize the intent of the initiative. During the first 18 months, the author (1) maintained notebooks on conversations and a record of correspondence and material, (2) prepared several reports for the foundation outlining project issues and progress, and (3) kept extensive field notes of visits to nonprofit organization SABI grantees. These data sources record the initial thinking behind the SOC and document the instrument's origins.

2. *Iterations of SOC.* The SOC has morphed through no fewer than six major iterations and several minor revisions of each iteration since the first draft. Together, they constitute a developmental record of the evolution of the instrument, with the first few drafts representing my initial thinking combined with that of several RWJF staff. Versions that followed were expanded by my reading of the literature and through field trials in community agencies.
3. *Consultation with expert evaluators.* Expert review was sought throughout the initial development of the SOC and included: (a) personal communication, (b) a RWJF convened meeting, (c) contracted consultation, and (d) a site visit by an evaluator other then the developer. The notes and correspondence from these consultations comprise a body of advice on validation of the SOC, as each of these is described more specifically below.

(a) *Personal communication* on the development was sought from eminent evaluators at both the 2000 and 2001 annual meetings of the American Evaluation Association (AEA). At the meetings informal advice was provided by John Stevenson, Paul Florin, Mary Ann Scherier, Vincent Francisco, and William Shadish all of whom provided references and encouragement in the initial development of what would emerge as SOC. I have notes and correspondence from these conversations.

(b) A *meeting* was convened by the RWJF program officer (Dr. Leviton) to consider issues in further developing SOC. Attending that meeting were Drs. John Stevenson, Paul Florin, David Chavez, David Hunter, Jackie Kaye, and Robin Miller. Dr. Hunter is leading a group at the Edna McConnell Clark Foundation working on issues of building organizational capacity. I have notes and correspondence relating to this meeting.

(c) *Ongoing consultation* on matters relating to SOC design, developing reliability, and validity was arranged by the RWJF. Paul Florin and John Stevenson of the University of Rhode Island and Vincent Francisco of Kansas University served in this capacity. Drs. Florin and Stevenson are evaluators affiliated with the Community Research and Services Team in the Department of Psychology. Dr. Francisco is affiliated with the Department of Human Development at Kansas University and is co-developer of the Community Tool Box, a comprehensive collection of tools and publications to enhance community organizational capacity made available through the internet. I have notes and written advice spanning a 12-month time from the consultants.

(d) *Site visit and pilot***.** Dr. Francisco was retained by the RWJF to assess the SOC development and administration. He reviewed the SOC construction and

administration instructions, after which he participated in administering the SOC at an agency, on May 10 and 11, 2001. Records of his correspondence, pilot test notes, suggestions, and his report to RWJF are sources of data for this study.

4. *Stakeholder feedback*. Two sets of stakeholder have provided feedback that contributes to the validation of the SOC; RWJF staff and staff in agencies that have been assessed using the instrument. Foundation responses to the SOC have been on-going and provide data about the formative process of both the development of the SOC and of the intervention it was developed to assess. This is a rich record because of complexity of issues, length of the formative process, and diversity of foundation stakeholder concerns. SOC profiles have been developed for 15 agencies to date. Our practice is to score an agency on the SOC and then meet with agency staff to present our observations and to obtain their feedback. Three agencies agreed to allow an in-depth assessment of their organizational capacity and the records of those transactions span periods of up to two years.

3.2. Data analysis

How these data will be used in various combinations to answer the three research questions is discussed next. The first question (a) addresses the methods for investigating the validity of the main features.

(a) *The initial development of the central features of organizational capacity and the maturity model*. Two sources of data present the initial thinking about organizational capacity from the conventional program perspective and from an evaluation perspective. One source is a "Needs Identification and Self-Assessment Tool" (SAT) developed by the SABI program coordinator as a means of negotiating specific agency interventions to

be funded by SABI. This instrument can be viewed as representing the conventional wisdom for such assessments since it was created by a process of cutting and pasting items from several existing instruments. The other source is the initial version of SOC developed in response to the evaluation needs of the project discussed in Chapter 2. The analysis will compare the central features of the two instruments and illustrate the differing views of capacity they promote. The product of this analysis will be an analysis of the similarities of the features of capacity addressed by both instruments and an explication of the differences produced by the two approaches in their initial application and the implications of those differences for SABI programming and evaluation

(b) Analysis of the revisions of the initial SOC based on field trials in SABI grantees and in nonprofit community agencies in Pittsburgh. The revisions of SOC resulting from pilot applications record the evolution of the initial concept as adapted or modified by practice. My plan is to compare the initial (July 23, 2000) and (February 2, 2003) latest versions and present the rationale for the changes. The result of this analysis is a description of why changes in the SOC were necessary including the rationale for and the benefit obtained from them.

(c) *Assessment of interrater reliability among the three SOC project staff*. The iterations of SOC provide evidence of its initial reliability. The early application of SOC was examined for evidence of reliability among staff using the construct of analysis of the number of agreements divided by the sum of agreements and disagreements suggested by Miles and Huberman (1984). My plan was to array each rater's scores on each attribute for each feature and sub-feature across several observations scored on the SOC. This analysis was intended to provide more help in explaining development of increasing or

decreasing reliability. Observations are made by each of three raters, the possible comparisons are:

(a) with (b),

(a) with (c), and

(b) with (c)

A six-by-six contingency table was used to analyze the extent to which scores of each pair of observers agree. Each of the SOC features was scored individually rather than in the aggregate for each organization. The cells of the contingency table represent the stage of development with a cell for instances in which one rater scored an attribute and the second rater did not. These were combined to give per cent agreement by features. An overall per cent agreement was not calculated as the features were unevenly developed and it would have been meaningless. The initial per cent agreements were used to make more precise definitions for each attribute, and clearer distinctions between them. Results should show improving reliability among the raters over time as the SOC was more precisely defined.

(d) *Feedback from RWJF staff and agency officials.* Judgments of experts are frequently used in the validation of instrument construct and content validity. Indicators of the validity of the SOC, then, are the responses of various experts and stakeholders affected or potentially affected by it. Responses from RWJF, consultants retained by RWJF to provide advice them on SOC development, and agencies assessed were analyzed for evidence of validity or lack thereof. These were considered the best available "experts" for determining the accuracy of the SOC for measuring organizational capacity to

undertake new of expanded projects. The next chapter presents the findings of these converging lines of evidence.

3.3. Data collection and use of the SOC

The SOC was designed as an observational instrument to assess organizational stage of maturity *at the time of the observation.* Raters were trained to follow a set of guidelines setting the conditions for use of the SOC. These general guidelines included pre-observation visit preparation, information sources, and time required to complete the SOC. These general guidelines are briefly outlined below.

Pre-observation preparation. All SOC observations were made on-site at the organizations being profiled. Raters were trained to acquire and study as much information about each organization as possible prior to going on site to make the sites visits as efficient as possible. An IRS 990 was reviewed for each organization to obtain an understanding of the general financial resource status of the organization. IRS 990s are usually available through database services or may be requested directly from the organization. If a 990 was not available, two or three annual financial reports would serve to orient the observer.

Organizational and staffing charts were requested in advance but not always available prior to a site visit. General information like age of the organization, size of the board, tenure of the executive would be collected if readily available and frequently some of this

information was available from foundation files. A series of board minutes would be reviewed to analyze board member participation. Information not available prior to arriving on-site was requested at the beginning of the visit.

Information sources. Observers interview an organization's executive, board members, and key staff members to gain an understanding of how the organization actually functions. The intent is to obtain information about structures, formal and informal systems, the extent of task specialization and role differentiation, and the specific nature of internal operations. Observers were instructed to identify attributes for each of SOC feature that indicate stages of maturation or the direction of change of a feature. Interviews would continue until no additional information was being obtained or until the observer had sufficient information to rate 12 or more attributes of each feature.

Time required. The time required to produce a SOC profile varies with the size and complexity of the organization being observed from a day to several days. For example, one SABI organization observed had been in existence for more than five years, but had only two full time staff, limited space and infrastructure, and a single core service. The SOC profile was produced in a matter of only a few hours. On the other extreme, a large Health Center with an annual operating budget of 11.5 million dollars, operating in several communities, and offering a range of medical and dental services, required two week-long visits to produce a SOC profile accurately.

4. VALIDITY OF THE SOC

4.1. Context for initial development of the SOC

On February 18, 2000, the Robert Wood Johnson Foundation (RWJF) held a meeting to consider a demonstration program initiative to build community agency capacity. Three groups of stakeholders attended the meeting: program staff, financial staff, and evaluation staff. Each stakeholder group was represented by three levels of responsibilities and experience: a vice-president, senior program or finance officers, and program or finance officers. The attendees represented, in the developer's experience, a diverse and impressive array of experience and expertise. The author was invited to attend the meeting as an evaluation and community organization consultant based on experience in both domains. The foundation wanted to consider a problem it was experiencing when making grants to small community organizations. Often, these organizations were unable to complete RWJF projects as proposed. This problem was thought to be especially acute with smaller community agencies and related to insufficient organizational capacity. A demonstration initiative was being considered to improve agency capacity to implement new or expanded activities or to sustain them after RWJF funding was completed. The contemplated demonstration was to consist of training and technical assistance (T&TA) for a period of three to six months with "catalyst" grants available to some of the grantees with which to make limited purchases to support their participation in the T&TA. The

interventions were estimated to cost $15,000 to $20,000 each. The staff envisioned an initiative of three to five years with as many as 15 to 20 grantees per year. The demonstration was named the Small Agency Building Initiative (SABI) and was planned to begin within 60 to 90 days of the meeting, about May 2000.

As a former small agency executive director, the author agreed that the initiative addressed a significant problem. As an evaluator, the author had concerns about the evaluability of the initiative as proposed, based on both experience and review of the evaluation utilization literature. The chief concerns were the lack of a formative evaluation stage and an accelerated time table. Often social interventions are not fully developed or implemented before they are evaluated, resulting in an unfair test of the intervention. In federal programs, Wholey and colleagues found premature evaluation so prevalent that they created "evaluability assessment" as a method for determining whether an intervention or program was sufficiently developed to support an evaluation (Horst et al., 1974; Rutman, 1980; Wholey, 1994). Lipsey concluded that the large majority of evaluation research designs amount to "black box" research because of poorly developed theory or causal models (Lipsey, 1988; Lipsey et al., 1985), or because the designs and measures are not of sufficient sensitivity to detect the intervention effects they purport to measure (Lipsey, 1990). Reichardt (1994) asserts that evaluations fail to detect effects because the interventions or programs are of insufficient strength to produce effects large enough to measure. So pervasive are failures to adequately develop implementation monitoring, intervention models, or evaluation methods, that the term

formative insufficiency was coined to describe the general failure to adequately prepare for outcome or summative evaluation (Leviton & Schuh, 1999).

For this study, the formative evaluation phase or formative stage is understood to include all of the front-end effort necessary to prepare social interventions adequately for successful summative or outcome evaluation (Wholey, 1994; Reicardt, 1994; Lipsey et al., 1985 & 1990; Sechrest et al., 1979; Tukey, 1976; Weiss, 1972; Suchman, 1970; Scriven, 1967). The accelerated timetable planned by the foundation implied that not enough time would be available for an adequate formative stage in which to develop an intervention, select grantee participants to receive the intervention, and develop appropriate evaluation measures to assess the effects of the intervention.

The absence of a sufficient formative stage has several potentially adverse implications for successful evaluation of social interventions, among which are:

A formative stage is frequently necessary in order to optimize an intervention (Scriven, 1967). Less-than-optimal interventions represent weakened or diluted implementations of what could possibly be strong ones. Evaluation of them results in ambiguous or null findings. Consequently, potentially sound interventions do not receive a fair test, owing to flawed implementation (Reichardt, 1994; Lipsey et al., 1985; Yeaton, 1985; Yeaton & Sechrest, 1981; Sechrest et al., 1979).

A formative stage is where intervention concepts are piloted to see how they will work in real-world settings (Suchman, 1970; Weiss, 1972). Refinement of intervention

components or concepts frequently takes the iterative form of implementation, assessment, refinement, and implementation. Failure to complete the refinement of concepts prior to evaluation can result in changes to interventions that make the designed evaluation irrelevant to the "morphed" intervention.

A formative stage is where measures to be employed are validated for use in evaluating the specific intervention. The sensitivity of the proposed methods and measures to detect intervention effects should be understood prior to employing them in an experimental or quasi-experimental evaluation (Lipsey, 1990). Reichardt (1994) observed that most reported social interventions would be able to produce only small to moderate results. Lipsey and his colleagues found that evaluation designs had the power to detect only large to moderate effects in as much as 60% to 90% of the evaluations they studied. On a practical level, the author has frequently heard program staff assert that they "know we make a difference" regardless of the ambiguous evaluation findings.

The SABI initiative needed a formative stage for work in all three of the areas. The agencies chosen to participate in SABI were likely to be varied in their organizational structures, the services they provided, and their developmental needs. Therefore, T&TA interventions that would be provided to these agencies could also be expected to vary. There was no way of planning what might be required to optimize the intervention before its development and implementation. The author recommended a two-year formative stage followed by a three-year summative evaluation stage. The formative stage, in which the intervention, management infrastructure, and evaluation methods and instruments would be developed, would consist of an iterative process of development, pilot implementation in a few organizations, and revision as necessary. This process was

intended to be consistent with Suchman's (1967) five-stage model of demonstration program development as updated by Carol Weiss (1972) and the author (see Figure 2.1 in Chapter 2) and consisting of:

I. Exploratory stage,
II. Pilot stage,
III. Model stage,
IV. Prototype stage
V. Institutional stage

Stages I and II in this model are the formative stages in Scriven's distinction between formative and summative evaluation in that they are preparatory to testing models in stage III of Suchman's model.

A summative evaluation stage (stage III in Suchman's construct) of three years to test the SABI demonstration model would follow. This recommendation received the support of the evaluation staff at the meeting but was not completely endorsed by the program and financial staffs, who questioned the need for such extensive formative work. The result was a compromise whereby the SABI initiative was limited to two years with a three-month startup. During the two years of the demonstration, organizations were to be selected to participate in one of three cohorts of five organizations each. Development of new instrumentation was necessary because the state of the art of evaluation of nonprofit capacity was inadequately developed. The rationale for developing new instrumentation for evaluating SABI was:

1. The planned SABI intervention for each organization was limited in scope and duration,

implying that the resulting effects in organizational capacity would be predictably small to moderate at best (Lipsey, 1990; Tukey, 1976; Reichardt, 1994; Yeaton & Sechrest, 1981). Measuring small to moderate effects requires sensitive methods and measures that can detect change in capacity. No instruments meeting these criteria for assessing organizational capacity were found.

2. The concept of organizational capacity was a large molar object inasmuch as it comprises a potentially complex array of components among which the mediating and moderating relationships are largely unverified (Cook & Campbell, 1979). Improving the evaluation of nonprofit organizational capacity required development of more complex evaluation methods to match the complexity of organizational capacity and interventions to improve it (Lipsey, 1988) and to advance our understanding beyond simple description to explanation (Shadish, Cook, & Campbell, 2002). The decision to construct new instrumentation, called Staging Organizational Capacity (SOC), for use in evaluating SABI included an implied obligation to establish its validity, for two important reasons. First, the chances of conducting an evaluation that is recognized as successful are greatly improved with measures and methods that have been validated (Cronbach & Meehl, 1955; Freeman, 1971). Lipsey et al. (1985) conclude that the use of unvalidated instruments is a major contributor to the failure of evaluation to produce other than ambiguous or null results.

Second, evidence of validity is essential for gaining acceptance of innovations on established practice or business-as-usual (Patton, 1978). The failure to achieve consensus on the recommendation for the nature and duration of a formative stage for SABI was an indication

that a need for innovation in methods or instrumentation was not shared equally among the three stakeholder groups: the evaluation, program, and fiscal staff. It was incumbent on the evaluators to demonstrate the efficacy of investing in development of the SOC. It was felt important to produce evidence that the SOC could yield *different* or *better* results than conventional practice for assessing nonprofit organizational capacity.

The remainder of this chapter presents evidence for the validity of SOC in four sections: The first section examines evidence that SOC produces results on the major dimensions of organizational capacity that are more comprehensive and detailed than those occurring with conventional approaches and instruments. The second section examines evidence that the SOC operationalizes or clarifies the meaning of organizational capacity, especially as it affects an organization's ability to implement expanded or new projects. This is an aspect of construct validation as used by Cronbach and Meehl (1955). The third section investigates evidence from three raters with different nonprofit organizations that the SOC produces similar results when different users apply it to the same organization. This dimension relates to reliability across users, which is necessary for validation of the instrument as a tool for use by other individuals and organizations. The fourth section presents further evidence of the SOC's construct validity and practical utility through reactions from experts and stakeholders as it was used in practical settings.

4.2. Comparison of the SOC and SAT instruments

Background. The SABI program was preparing to select an initial cohort of grantees in fall 2000. Early iterations of the SOC had been circulated at that time. However, the SOC was an original approach, synthesizing knowledge from such diverse sources as cognitive psychology to systems theory, and evaluation to organizational development. This synthesis was well grounded in social science and experience but it had not been validated as more relevant than conventional methods of assessing the capacity of nonprofit community organizations. The SABI program staff was reluctant to rely on the SOC profiles for an initial assessment of grantee capacity, preferring to adopt conventional approaches employed by other foundations. They wanted a self-assessment approach by which grantees would set priorities on the type of T&TA they wanted. This, they felt, would allow the SABI to be more efficient by combining group training in a central location with on-site assistance to reinforce the training. The author had presented evidence to support the efficacy of targeting the SABI intervention to the stage of development of the organization receiving it. Stage-appropriate interventions tend to be stronger and more successful than are more general or one-size-fits-all interventions (Kazanjian & Drazin, 1990; Yeaton & Sechrest, 1981). The evaluation staff was convinced of the value of the SOC and committed to its development and use. Negotiations failed to produce a consensus on use of a single instrument, and it was understood that the program staff would adopt or develop a conventional tool to be administered concurrently with the SOC in initial assessment for the first cohort grantees. Therefore, two alternative instruments were being developed at the Robert Wood Johnson Foundation: the Needs Identification and Self Assessment Tool (SAT) being constructed

by the program side of SABI and the Staging Organizational Capacity (SOC) instrument under development by the evaluation side. The SAT was constructed in a two week period in November 2000 and comprised items taken from several instruments used by other foundations for assessing organizational capacity. These included the National Endowment for the Arts' Organizational Self-Assessment Tool, Local Initiative Support Corporation's Capacity Building Survey, and an Irvine Foundation Assessment Tool. None of these instruments had reported assessment information on validity or reliability. However, even had these instruments been validated, there could be little or no carryover to the SAT. Established reliability and validity could be expected to change when applied in novel or adulterated ways (Shadish et al., 1981). The SAT, comprising items from several existing instruments, was a new and untested instrument, having only face validity based on the previous instruments and SABI developers' expertise.

The SOC had been under development for six months and involved a set of three progressive steps. The first step was to develop a set of features that were postulated to change as organizations matured. These were applied retroactively to organizations with data on file as a result of previous work. For example, the author assessed 26 small community organizations for the William Penn Foundation's anti-violence program and had extensive experience with one demonstration program funded by the Staunton Farm Foundation. An iterative process was applied, consisting of (1) development of features from the literature, followed by (2) application to a case in the files, followed by (3) revision as indicated by the application, and finally (4) application of the revised SOC to another case.

Evidence that the SOC had the potential of moving beyond description of an organization's capacity to explanation for that status was emerging at this very early developmental period. An illustrative example of the explanatory capability of the SOC occurred when it was administered *ex post facto* to the Staunton Farm project. The project was a collaborative effort between the University of Pittsburgh and a new nonprofit community organization. The foundation was interested in supporting the new organization as well as piloting the demonstration. The project did not accomplish its objectives to implement the demonstration intervention as planned and failed to perform at the levels proposed. From the university perspective, the project was a failure because the community organization was unable to fulfill its agreed upon responsibilities. The SOC profile revealed that the organization did not possess sufficient capacity to operate on equal terms with the university. The staff was cross-subsidized for all administrative functions (Estelle, 1983). In other words, project funds earmarked to pay staff to deliver services were their only support. However, staff needed to perform additional organizational duties not supported from other sources like book-keeping, proposal writing, and other fund raising-activities. The SOC provided a framework for analyzing and explaining this phenomenon in a more objective manner. Viewed from this new perspective, the project was not a complete failure. The cross-subsidization contributed to sustaining the organization, and it has remained a viable part of the community. Had we had the insight of the SOC perspective, the project could easily have been designed to better accommodate both the demonstration project needs of the University and the capacity building needs of the community organization of concern to the Staunton Farms Foundation.

The second step in the evolution of the SOC was to apply the instrument in the field. Taking this step required recruiting volunteer organizations that would agree to participate in having their SOC profile completed. The first organizations were ones with which the author had worked in some capacity in the past. Organizations in Pittsburgh and in the Washington, D. C., area were selected to reduce travel time and expenses. Five organizations agreed to participate in exchange for receiving an orientation to the SOC, the resulting profiles, and assistance in understanding the implications of any issues that might arise as a result of the profiles. By agreement, the presentation of each profile was two-way, with the organization receiving the observations of an outside evaluation professional and the evaluator receiving feedback about the accuracy and utility of the process. Each of the organizations in which the SOC was piloted confirmed the accuracy and relevance of the profile, and each requested additional technical assistance based on the profile developed.

One organization profiled was a community mental health center with an annual operating budget of approximately $11.5 million. The executive director requested a follow-up to the presentation of the profile for the purpose of "discussing discrepancies with the SOC profile." With feelings of concern over what the profile had missed, the author and his colleagues reminded themselves that encountering some disagreement with the reported profile was the purpose of the pilot and that some error was inevitable. Surprisingly, the executive director did not disagree with the profile. Rather, she wanted

to discuss ways to reduce the variance between her preconceived notion of the organization's maturity and the SOC profile. Organizations participating in the pilot test agreed, without exception, that the resulting descriptions of their organizations were accurate. Two major lessons were learned by the "real world" pilot test; (1) maturity on features was not necessarily related to size of the organization, and (2) accurate, meaningful profiles of organizational maturation could be completed rather quickly. The organizations in which the pilots were conducted ranged in size, as indicated by their annual operating budgets, from $300,000 to $11,500,000. Their organizational ages ranged from two years to 22 years.

Initial profiles were based on varying amounts of exposure to the organizations (see Chapter 3). The first one was completed through a series of weekly meetings of one to three hours' duration over the course of three months. Another was based on a two-day intensive on-site visit.

The third developmental step of SOC applied drafts of the two instruments, the SOC and the SAT, to five SABI grantees in November and December 2000. The data available for application of both instruments were identical since they were completed during interviews attended by both program and evaluation staff. A necessary first step in validating the SOC was to demonstrate to the program staff and the evaluation staff that the SOC and SAT addressed comparable domains, but that the SOC addressed organizational capacity more clearly and in greater depth than the SAT. In the remainder

of this section, the SAT is considered to represent the normal or usual method of assessing organizational capacity, since it was derived from several exemplars for assessing capacity from the program staff perspective and was employed for SABI. The SOC is compared to the SAT to document that it addressed central dimensions of organizational capacity, and to describe differences in form and function between the two approaches for assessing organizational capacity. The next section describes the two instruments and their development, followed by a discussion of their similarities and differences, and concluding with an analysis of the effect of those differences on the information each provides.

4.2.1. The Self Assessment Tool (SAT) (Draft of November 20, 2000.)

The SAT, for use in assessing grantee organizational capacity technical assistance (TA) needs and priorities, represents a conventional approach to organizational assessment. Its construction, based primarily on the literature and experience of other major foundations, used parts of existing instruments without attention to validity other than the stated purpose of assessing organizational capacity for their development. The purpose of the SAT was to identify grantee capacity development needs so that T & TA could be targeted to those needs that grantees shared in common. The SAT is intended to be completed by grantee staff, who were asked by a foundation program manager to rate attributes of each of 11 central features on a four-point scales as:

(1) Well developed,

(2) Adequately developed,

(3) Needs development, or

(4) Not applicable.

The eleven features of the SAT rated were:

- *Organizational Purpose.* The organization's mission and vision statements and a written history that are reviewed, understood, and agreed upon by key stakeholders.
- *Governance.* Attributes of the board of directors and their policies and procedures.
- *Planning.* Comprehensive multi-year formal plan that is updated annually.
- *Legal*. Access to counsel, the existence of formal personnel and grievance policies, compliance with reporting and legal requirements, and the existence of outstanding lawsuits.
- *External Environment.* The understanding and knowledge of the community in which an organization operates.
- *Programs.* Program planning, communication of vision, method of evaluation, and long-term planning.
- *Staff/Communications/Decision-making.* Reporting relationships, job descriptions, administrative leadership, staff and board communications, planning, staff experience and formal human resources functions like personnel policies and performance reviews, salary scales, staff morale, and office equipment.
- *Public/Community Relations.* Clarity of organizational image communicated to the public, the organizational chart, handling of complaints and queries from the public, annual public relations plan, effectiveness in representing the organization,

size of staff and budget, and the organization's publications.

- *Financial Management.* The budgeting process, computerized accounting systems, system for cash flow projections, ability to pay accounts within 30 days, system of internal controls, payroll tax deposits, lines of credit, investment returns, staff size and experience.
- *Computer Systems and Hardware.* Computer support in various functional areas like financial management and planning, computer training for staff, software systems, and hardware acquisitions.
- *Fund raising.* Process for setting annual goals; mechanisms for staff and board input; size and experience of staff; grant writing expertise; levels of funding from government, nonprofit, profit, and individual sources; mechanisms monitoring and expanding donor base, and budget for fund-raising expenses.

In addition, grantees were asked to prioritize the "three most pressing technical assistance needs" as identified by staff. It was felt that grantees would identify those features for which they were deficient and would identify their highest priority needs as being the most pressing. These would then be analyzed by the program staff to identify clusters of need that would then be targeted to obtain maximum benefit from the limited T & TA resources. The intervention could combine centralized training with on-site individualized technical assistance for greater economy.

4.2.2. The Staging Organizational Capacity (SOC 1), (Draft of November 2, 2000.)

The SOC 1 is a maturity model developed to assess an organization's stage of development on key features that affect its ability to implement new or expanded

community innovations (see Appendix B). Grounded on a wide range of organizational, systems, and management research and associated literature, it is scored by an experienced organizational development observer. Aspects of six central organizational attributes are rated on five stages of development or maturity ranging from "least developed" to "most developed." The six SOC 1 features were:

Governance. Legal status of organization and associated documents, board composition and functioning, and the relationship between board and staff.

Service Systems. Horizontal and vertical structure of the core services.

Service or Programs Skill Development. Skill and experience of staff in core service roles.

Role Function and Specialization. Degree to which an organization has differentiated its service and administrative functions; degree to which tasks are specialized by degree of skill or training they require.

Administrative Skill Development. Skills and experience of the administrative staff.

Financial Maturity. Organization's funding history and its ability to generate funds with which to sustain itself.

These features were chosen because they were both (a) consistent with the program staff's features of interest and (b) appeared to change over time based on experience and evidence in the literature that they vary with organizational maturity, growth, age, or life-cycle.

The SOC 1 produced a profile of organizational stage of development on each of the six features that is based on the ratings of an outside observer who has extensive experience in studying organizations and their behavior. The rationale for the SOC 1 was based on the organizational life-cycle literature, which suggests that organizations have a general pattern of maturation in common. Cognitive research suggests that intuitive recognition of underlying patterns is a characteristic developed by experts in a given domain. Putting these maxims together, it seems reasonable to expect that SOC 1 observers with considerable organizational experience would produce comparable profiles based on underlying patterns of organizational maturity, thereby, conferring on the SOC 1 a high degree of reliability and sensitivity based on observer expertise.

The SAT and the SOC overlap considerably in the dimensions of organizational capacity they measure. Yet, they tend to produce different results with the SAT focused on formal development of policies and procedures and the SOC 1 more focused on the dynamics of organizational growth and maturation. Explaining how and why this is so requires a closer examination of the similarities and differences between the two instruments and is the subject of the next section.

4.2.2.1. Similarities and differences between SOC 1 and SAT.

Table 4.1, below, arrays the central features of the SAT and the SOC 1, and aligns them by functional substance as closely as the two approaches permit. The numbered and italicized items

in Table 4.1 indicate the central features of organizational capacity as presented in the description in the previous section. Selected attributes of each feature are identified and listed below the feature they describe. Attributes of a feature are not italicized. The SAT contains eleven features and the SOC 1 contains six. These features have been aligned to reflect functional aspects of capacity treated by each instrument. One will notice considerable functional overlap in the features of capacity addressed by both instruments. Only the SOC feature (6) Financial Maturity does not have some functional overlap in the SAT. Each instrument defines governance differently. For example, the SOC defines governance more broadly than the SAT as can be seen in table 4.1, which shows that the SOC feature (1) *Governance* functionally relates to attributes of three SAT features; (1)*Organizational purpose, (2) Governance, and (3) Legal.* The attributes of governance for the SAT reveal a focus on formal development of documents and policies like "written mission statement" or board "job descriptions." The SOC 1 focus on governance is not on formal development but, rather, on organizational behavior like how the board organizes itself or how it makes decisions.
The SOC 1 did not assess mission statements since they have not been demonstrated to change in systematic ways as organizations mature and do not directly alter an organization's capacity to expand their current services or start new ones.

The SAT feature (3) *Legal* continues the pattern of placing emphasis on the "adequacy" of formal development of policies and procedures and has no direct analog in the SOC. Instead, the

Table 4.1:

Central Features of SAT and SOC

Self Assessment Tool (SAT) November, 2000	Staging Organizational Capacity (SOC 1) November, 2000
1. *Organizational purpose* Written mission/vision statements & organization history	1. *Governance* Board of Directors Selection and characteristics Bylaws Meetings and attendance Functioning/roles Organization of Board Board decision-making Charter and Licenses
2. *Governance* (Board of Directors) Job descriptions By-laws Meetings and attendance Functioning/roles Involvement of staff	
3. *Legal* Counsel Insurance Policies and Procedures	
4. *External Environment* Local social & economic climate Knowledge of community & region	2. Appropriateness and adequacy of *Service Program Skills*
5. *Programs* Provision for maintaining continuity Costing program elements Evaluation	3. *Service Systems Maturity* Vertical/horizontal differentiation Service niche *4. Role Function and Specialization* Service and Administrative differentiation
6. *Staff/ Communications/ Decision-making* Organizational chart Staff & board communication Staff size appropriate Job descriptions Salary scale Office equipment	*4. Role Function and Specialization* Service and Administrative differentiation

7. *Planning* Comprehensive written organizational plan	*4. Role Function and Specialization* Service and Administrative differentiation
8. *Public/Community Relations* Effectiveness in representing the organization Mechanism for handling public communications Organization's publications PR budget	5. *Administrative Skill Development* Staff skills & experience
9. *Financial Management* Budgeting system Accounting system Written cost controls Annual audit Payroll tax deposits Debt management Line of credit Returns on investments	
10. *Computer Systems and Operations* Computer training for staff Software systems Hardware acquisitions	
11. *Fundraising* Government grants Development software Donor information system Fundraising budget Grant writing expertise Development research capacity Development staff	
	6. Financial Resources Maturity Funding history and stability

SOC reviews legal or conditional constraints contained in organizational charters and licenses since they can and, frequently do, directly limit or constrain capacity. For example, an inpatient substance abuse treatment center that could potentially benefit from expanding its services may be constrained by the number of beds permitted by its license.

The SAT feature (4) *External Environment* has no direct SOC 1 analog because many SOC organizational features can be shown to contribute to environmental relationships, including the board, the executive director and key staff, and core service providers. Also, the SOC 1 recognizes that the robustness or strength of an organization's programs or services as an indicator of its environmental relationships. In other words, if programs are not meeting a need they will not long be in business.

The SAT feature (5) *Programs* is most analogous to the SOC 1 features (3) *Service System Maturity* and *(4) Role Function and Specialization.* The SOC 1 examines the maturity of programs or services of nonprofit organizations in terms of their structural characteristics. For example, taller and wider structures are usually found in more mature organizations. How organizations cost their services also differs with stage of maturity and is, in part, a function of administrative/service differentiation and task specialization. The SAT assesses program attributes relating to sustainability and evaluation and the SOC 1 assesses how services are structured vertically and horizontally. The SOC parses the SAT attributes to other features. For example, the SAT attribute "Costing program elements" is accommodated by the SOC 1 as a function of administrative differentiation.

SAT attribute "Maintaining continuity" is parsed as attributes of the SOC's assessment of (6) *Financial Maturity.*

The SAT feature (6) *Staff/Communications/Decision-making* is a conglomerate feature analogous to the SOC 1 feature (4) *Role Function and Specialization,* which seeks to describe the degree to which organizational roles are differentiated, or separated, and the degree to which those roles have higher level functions supported. For example, in small, less-developed nonprofit organizations, one can frequently find an executive director who is responsible for planning, reporting, human relations functions, and fund-raising. Such a director may be highly experienced and well motivated. But that director will be hard-pressed to perform all of those functions consistently at a high level. More mature organizations will support the executive director's job and specialize those various roles to two or three deputies, thereby improving performance. It is not unusual to see small organizations possessing a bookkeeper in addition to an executive director. Typically, the person holding that job has a variety of responsibilities, some routine like getting the payroll out and filing tax reports, and some higher level tasks like budgeting and forecasting. In this case, greater productivity can be achieved by differentiating the routine tasks from the higher order tasks. In more mature organizations we find more experienced financial staff freed from the routine tasks in order to attend to higher order tasks by the support of less experienced and less costly staff clerical and bookkeeping staff.

SAT features (7) *Planning,* (8) *Public/Community Relations*, (9) *Financial Management*, (10) *Computer Systems and Operations, and (11) Fund-raising* are all internal operations that may be components of an organization's administrative infrastructure. The SAT implies that all organizations have, or should have, these features, and that they can be assessed as "well

developed, adequately developed, needs to be developed, or not applicable." The SOC 1 treats characteristics like public relations and communications as internal operations that may or may not exist within a given organization and for which the stage of maturity of each depends on the maturity of the SOC features (4) *Role Function and Specialization* and (5) *Administrative Skill Development.* Not every organization will have all of the SAT features like computer systems or public relations functions. But all organizations will have internal operations that can be assessed in terms of functional differentiation, task specialization, and experience and skills of staff. As discussed above with planning, internal operations tend to be put together uniquely by organizations using whatever resources may be available. Internal operations in nonprofit organizations may be pieced together or structured in a variety of ways; integrated or dispersed, centralized or decentralized, formal or informal. Therefore, they appear unique with each organization making them difficult to describe systematically. However, the productivity and maturity of all internal operations can be expressed in terms of the SOC features (4) *Role Function and Specialization* and (5) *Administrative Skill Development.*

One feature on Table 4.1 remains to be described, the SOC feature (6) *Financial Maturity,* which stands alone as having no functional analog on the SAT. Nonprofit organizations appear to systematically mature in the manner in which they obtain their funding from unstable sources to more stable sources, from short-term to long-term funding, and from inadequate cost recovery to more reasonable cost recovery. Considerable evidence supporting the importance of financial maturity is derived from the recently emerging research on financial vulnerability made possible with the availability of the Internal Revenue Service form 990 reports required of nonprofit

organizations. This research suggests that financial maturity is a function of the stability and robustness of organizational income. Stability is used in the sense that operating margins are positive and robustness in the sense that organizations possess diverse sources of income (Tuckman & Chang, 1991; Greenlee & Trussel, 2000; Ritchie & Kolodinsky, 2003). This research has achieved good validity in predicting nonprofit organizational failure. The SOC applies indicators of robustness or vulnerability in assessing organizational funding history to identify trends in either direction as indicators of capacity to expand or take on additional work.

Analysis

The two approaches to assessing organizational capacity identified different capacity-related issues during site visits to the five first cohort SABI grantees. The priorities or "needs" identified through use of the SAT were similar in all five agencies of the cohort. All five agencies placed high priority on some form of "marketing and public relations." In addition, four of them also indicated that technical assistance in fund-raising was a priority. These two clusters justified an emphasis on group training in the areas of communications and fund raising. Less focused needs, like computer technology and database development, were left for individual technical assistance or were not addressed.

Use of the SOC produced a more varied picture that did not emerge from the Self Assessment Tool. For example, the *Agency #2* and the *Agency #5* were both assessed as agencies in transition, either to a higher developmental level or expansion at the same

level. Both cited problems resulting from sustained growth of clients as threats to maintenance of service quality. Both agencies were sufficiently extended that their capacity to handle more clients or new programs was limited. For both agencies, increases in capacity necessary to sustain growth had been obtained through increases in staff productivity (i.e., working longer hours and neglecting some tasks in favor of others). There clearly was not much more capacity to be obtained in this fashion for either agency.

The *Agency #4* was making a significant jump from one level of operations to another. The operating budget of $339,000 represented a 62% increase over that of each of the previous two years. Making such growth does not automatically mean that the agency increases its capacity. In fact, the Center reported that it expected to experience start-up problems with its new project. Start-up problems are a significant indicator that an agency does not have the reserve capacity to implement new or expanded endeavors.

The A SABI funded Health Center was found to have highly developed core services but an administrative capacity and governance structure at a similar level of development as that of the much smaller *Agency #*5. Both organizations were programs in much larger parent organizations with governance provided by advisory boards. Both had small part-time administrative staff. Both organizations had the advantage of being able to approach their parent organizations for additional support in the event of crisis or emergency.

The overall SOC assessment of the first cohort identified four agencies experiencing transitional stress, two with stress from growth or expansion and two, with stress from environmental pressures. The fifth agency appeared stable and occupied a niche in which it had no competition and considerable stability. Capacity building T&TA could target improving productivity or infrastructure in the growth organizations for maximum effect. The two organizations experiencing environmental stress would benefit from assistance in improving their support development capacity. The fifth agency might best benefit from assistance in differentiating its administrative infrastructure or its governance functions to build capacity as indicated by the lack of development of administrative/service differentiation by the SOC.

The SAT described *Agency #*2 as placing high priority on "computerization." The SOC assessment revealed that senior staff were over-extended, in part because of having to do repetitive tasks and, in part, because of a need to do a mix of lower and higher level tasks. "Computerization" was their view of a method of improving senior staff productivity by freeing them from the need to work on lower level and repetitive tasks. Computerization was interpreted with the SOC 1 as an expression of a symptom of an underlying problem rather than a solution. Several factors explain the differences in the SAT assessments and the SOC 1 profiles of the first cohort of SABI grantees;

1. The SAT assessed organizational priorities in terms of adequacy of the formal processes and functions (e.g., job descriptions, plans, procedures), while the SOC used observational methods considered degree of formality as an indicator of stage of development.

2. The SAT rating method was designed as a *self-report instrument* to elicit a prioritization of technical assistance "needs," while the SOC employed *observational* methods to profile the pattern of development of organizational features.
3. The SOC 1 emphasizes on developmental dynamics while the SAT focuses more on static indicators.
4. The SAT focuses on how organizations present themselves to potential supporters while the SOC is concerned with how organizations actually get things done.

These four points are considered in greater detail below.

The SAT emphasis on adequacy of formal processes like long range planning, personnel policy manuals, and well crafted mission and vision statements was consistent with the goal of securing agreement on general substantive areas for which training and technical assistance could be provided. This goal contrasted with the SOC goal of assessing organizational capacity to undertake new or expanded projects. Not surprisingly, the results of applying the two instruments differed fundamentally. The SAT asked each agency to record its self-assessment of its T&TA needs at the end of each section (Organizational Purpose, Governance, etc.). There were nine sections for which T&TA was plausible, and there were five organizations in the first cohort, making possible a total of 45 individual responses. Organizations identified T&TA needs in 42 of the 45 (93%) spaces provided. The nature of the instrument's design and the process of its administration appeared to demand a response.

The SAT asked each organization to identify its T&TA needs. The use of the word need carries with it a connotation that the organizations were deficient in some quality necessary for good, or improved, performance. The SAT emphasized formal development of plans, policies, and procedures. The effect of this focus is to place higher value on more formally developed organizational features and to devalue informal ones. The SOC, on the other hand, recognizes the degree of formality or informality of functions or activities as indications of stages of maturity. The SOC 1 does not imply that more formal systems are better than informal ones. In fact, there are sound reasons for some organizations to eschew or reject more formal development. Paul Light (1988), for example, observes that small horizontally structured organizations are not necessarily less productive than larger more vertically structures ones, but are more often innovative. Systems theory suggests that informal organizations are more responsive to their environments (Blau & Scott, 1962). To illustrate the difference in the two approaches consider the following example. The SAT asked organizations to describe their "comprehensive, multi-year organizational plan in place and in use" on a scale from well developed to needs development. A small community organization in the first SABI cohort responded that it did not have such a plan. From the SOC perspective, not having a written plan did not mean that it had no plan. The board and staff spent many hours informally discussing how to overcome a projected budget short-fall in the current year. The organization reported experiencing similar problems most years. For SOC 1 purposes the planning process was informal but had been effective in the past. The planning was short-term as opposed to multi-year because of the immediate need to secure funding for their program for the remainder of the year. Both board members and

staff who were interviewed were in agreement about the necessary course of action. They did have a plan but it was an informal one.

By describing organizational features in terms of stage of development, the SOC seeks to determine the dynamics of organizational capacity. Features are not assumed to develop in only one direction. They may mature to a more formal or structured stage or they may regress to a less mature stage to be more informal or less structured. An example of this regression was found in a SABI community-based addiction treatment agency. When first visited the organization had recently begun to develop a more highly differentiated administrative infrastructure by adding a second full-time person to handle financial and planning functions. The organization had accumulated reserves equal to its six-month operating budget and was looking to expand its facility to accommodate the administrative growth and permit additional growth in service activities. At the same time the board of directors had declined from twelve members to four individuals. Meeting times and agendas had become highly informal and irregular. In the space of 30 months the organization appeared less robust, exhibiting signs of vulnerability. It had been certified to provide services supported from a new funding stream, but income generated from program services was not meeting the expenses. Financial and equity reserves were being consumed to meet the income short-fall and the organization was in crisis mode to get back to stability. This type of ebb and flow in maturation is not uncommon.

Summary.

1. Typical measures, such as the SAT, are developed to identify perceived "needs" and tend to emphasize formal attributes of organizations (bylaws, written plans, mission and vision statements, job descriptions, organizational charts, and annual reports). These self-identified "needs" are viewed from the perspective of participation in T & TA to more formally or "adequately" develop them. The SOC employs a maturity model to identify the stage of development of key organizational features. The SOC recognizes that how an organization represents itself to the world may not reflect actual maturity. Charters and licenses are used to determine mandated limitations on capacity, bylaws are used to determine how formally or informally they are followed, and planning and other organizational functions are used as indicators of stage of maturity.
2. The SOC 1 addressed nearly all of the topics addressed by the SAT, but often from a different perspective to focus on how the organization operates and on implications for capacity to undertake new or expanded projects.
3. The SAT uses organizational self-assessment to identify T&TA needs. The resulting assessment of needs is based on staff perceptions of solutions to problems they are experiencing. Frequently, these perceptions are symptoms of an underlying or more fundamental dysfunction. The SOC assesses stage of development of features of capacity based on the judgment of trained and highly skilled external observers. The resulting profile is more likely to address an underlying problem than a symptom of the problem.
4. The SOC 1 deals with the dynamic, inter-related nature of development of organizational capacity and provides a basis for targeting T & TA to specific attributes of each

organization. The SAT tends to look at a static picture from the perspective of a typical stage three of development resulting in a "one size fits all" intervention design.

4.3. Validity of SOC features and sub-features

Validity of revisions of the initial version of the SOC are analyzed in this section. Based on field trials with five SABI grantees and five additional nonprofit community agencies in Pittsburgh, these revisions document changes in the initial SOC conception of the maturity model as it evolved through a series of pilot tests. This evolution represents an important step in the validation of the SOC made necessary because of its unique and innovative nature. The concept of organizational capacity is generally viewed in molar terms (please refer to the discussion in Chapter 2) the parts of which are varied and vaguely defined. Previous work in this area has largely been reported as "lessons learned" or "best practices," indicating that development of a conceptual framework of organizational capacity has not advanced to the point of generating comparative data and systematic accumulation of knowledge about organizational capacity. This youthful state of development of the concept of capacity complicates attempts to evaluate interventions designed to influence programmatic outcomes of capacity-building efforts. The planning and conceptual efforts, both in terms of the interventions and of evaluation methods and measures used to assess them, represent our best judgments of the characteristics required to ensure success.

SOC 1 innovated on conventional treatments of capacity by synthesizing knowledge from several diverse literatures thought to be relevant for understanding the concept of

organizational capacity for undertaking new or expanded initiatives (see Chapters 1 and 2) and by using a maturity model that focused on dynamic rather than static characteristics. An iterative process of testing and revising the SOC in real-world conditions was undertaken as the second step in the process of developing and validating the instrument a step made necessary by the complex nature of organizations and the lack of validated theory to determine the features and attributes distinguishing one stage of development from another. Comparing the changes resulting from piloting iterations of the SOC in the field documents this step of validating the instrument and approach to measuring organizational capacity.

The rationale for these documented changes are analyzed and discussed below. Table 4.2 presents the features of both the initial SOC (SOC 1) and the final version as of this writing (SOC 11). As might be inferred from the version numbers, there have been 11 major iterations of the SOC recorded in the first three years of its development. The table reveals that the overall number of features has remained basically the same, but they have been realigned and sub-features have been added. Revisions were made based on several experiences:

1. Indicators of change from prior research did not fit the maturity model stages well.
2. Similar terms taken from different literatures were confusing to raters or unclear in application.
3. The results obtained were static and not producing information about the dynamics of change consistent with understanding organizational maturation that affected implementation or expansion of projects.

4. The results obtained were not directly related to organizational capacity (e.g. leadership).

These observations are considered below, with examples and discussion of how SOC 11 improves on the validity of SOC 1. Table 4.2 compares the features of SOC 1 with those of SOC 11. A discussion of the changes and the rationale for them follows.

Governance (feature 1 on Table 4.2) is defined broadly as the organization's legal status and policy decision-making behavior for both versions of the SOC. There is evidence that attributes of this feature change as organizations mature. However, nothing in the literature established an easy fit of these changes to the SOC maturity model approach. Therefore, SOC 1 was constructed based on general inferences from the literature to be verified or modified as the instrument was piloted.

In the field trials, it proved to be difficult to determine the direction in which this feature might be changing or be expected to change using the stage of development approach. However, a number of indicators of change in governance are described in the literature but had not been included because they could not be associated with a given stage of development. As an enhancement to SOC 11, the sub-feature "uncalibrated indicators of change" (1.A.) was added to accommodate predictors of the direction of change in governance that are independent of stage of development. For example, boards of directors that are expanding beyond their initial level are typically predictive of greater maturation from one stage to the next, whereas boards that are decreasing in size due to attrition usually indicate regression to earlier levels of maturation. Similarly, active

participation by most board members indicates interest and commitment necessary for development, while active participation of only a few key members usually suggests regression to an earlier level of maturation. The term "uncalibrated" indicator is intended to convey that these particular indicators are not related to specific stages of the model, but can indicate the direction of expected change in organizations at all stages.

In applications of the SOC, the uncalibrated indicators of change appeared to provide important behavioral information concerning the dynamics of change in the development of governance in nonprofit organizations and tended to draw attention to the importance of the dynamics of change in assessing organizational capacity.

Financial Resources (feature 2 on Table 4.2) maturity relates to the dynamics of funding streams of nonprofit organizations. In general, mature organizations tend to have more stable and diverse funding streams than do less mature organizations. The SOC was developed to characterize this feature on a continuum from highly unstable and unpredictable to highly stable and predictable. In the field, these stages have proved to be informative and accurate but did not directly address the issue of sustainability. Sustainability is understood in the nonprofit world as the likelihood that an intervention (or organization) will be continued after grant support is no longer available. The SABI, in particular, and the RWJF more generally, had indicated that "sustainability" of grantees and of grantee interventions was a primary concern. As indicated in Chapter 2, much has been written about sustainability, but little research is reported on predictors of

success in sustaining organizations or their interventions. However, some research is emerging on probability of nonprofit organization failure based on analysis of IRS 990 data. Tuckman and Chang (1991) identified variables predictive of nonprofit organizational failure. Their work has been replicated and extended (Greenlee& Trussel, 2000; Froelich & Knoepfle, 2000; Ritchie & Kolodinsky, 2003), and has added to our understanding of the potential robustness of nonprofits. These variables are reported to have good predictive validity. The dimension of financial vulnerability (2.A.) was added to the financial resources feature to more directly address the issue of sustainability. The synthesis of vulnerability research was the only addition to SOC 11 for this feature.

Service System Maturity, Feature 3**,** of the SOC 1 was intended to produce an assessment of the structure of each organization's core services. Applying the logic of organizational studies like that of Anthony Downs (1967) led to the hypothesis that taller organizational arrangements would tend to be more mature than shorter ones, and that they would also tend to be more formal and more highly specialized in role and task differentiation. At the same time, feature 4, *Role Specialization & Function,* of SOC 1 was based on the nonprofit and organizational life cycle literatures suggesting that administrative differentiation (and efficiency) is related to maturation (Letts et al., 1999; Light, 1998; Quinn & Cameron, 1983). It became clear in pilot applications that the two features were developmentally related to organizational maturity and were sub-features of organizational maturity. The chart reveals this redefinition on the SOC 11 as features 3 A and B under *Organizational Maturity. Role Specialization & Function*, feature 4, of SOC 1, was eliminated as a feature in SOC 11. Instead, Administrative Differentiation was recast as a sub-feature of *Organizational Maturity*. *Administrative Skill Development* and *Service Skill*

Development, SOC features 5 and 6, were developed in an attempt to integrate knowledge from cognitive psychology, organization development, and systems theory. The literature indicates that expertise is developed through experience and training (Brenner, 1984; Voss & Post, 1988;

Table 4.2:

Features of SOC 1 and SOC 11

SOC 1 Features	**SOC 11 Features**	**Description of difference**
1. Governance	1. Governance A. Uncalibrated indicators of change	Added to SOC 11
2. Financial Resources	2. Financial Resources A. Financial Vulnerability	Added to SOC 11
3. Service System maturity	3. Organizational maturity A. Administrative Differentiation B. Vertical & horizontal Differentiation	Redefined for greater precision as organizational maturity in SOC 11.
4. Role Specialization & Function (Administrative capacity)		Deleted on SOC 11 and incorporated as sub-feature in both internal operations and core services on SOC 11.
5. Administrative Skill Development	4. Internal operations A. Financial systems B. Management Information systems C. Communications D. Human Resource Systems E. Development	Redefined Administrative Skill Development as sub-features of internal operations in SOC 11.
6. Service Skill Development	5. Core Services (each service rated separately) A. Infrastructure B. Technology C. Role differentiation D. Task specialization E. Expertise of key staff 1. Domain expertise 2. Functional expertise	Redefined Service Skill Development as sub-features of each core services in SOC 11.

Ericcson & Smith, 1991; Bereiter & Scardamalia, 1993). In fact, Dreyfus, Dreyfus, and Athanasiou (1986) developed a five-stage maturity model to describe the development of expertise (although it was not yet called a maturity model at the time). Elements of their model of expertise have been incorporated into the SOC.

Cognitive psychology has demonstrated that expertise is highly domain specific. That is, expertise in one domain does not necessarily translate into expertise in other domains (Chase & Simon, 1973; Glaser & Chi, 1988). My experience with small community organizations was that many of the leading new organizations had developed a high level of expertise in their field, but then found themselves less prepared to function in the new domain of administration of a nonprofit organization. This experience is consistent with research findings that infrastructure, technology, and task specialization can determine the degree to which individuals are able to function at their highest skill levels (Vinzant & Vinzant, 1996). In pilot applications it became clear that the fundamental principles applied equally to what had been termed administrative skill development and what had been termed service skill development. The concept of skill development or expertise was incorporated into SOC 11 as sub-features of each *Core service* (see 5C Role Differentiation, 5D Task Specialization and, 5E Expertise), the rationale being that Expertise is an individual attribute that can be measured on the Dreyfus, Dreyfus, and Athanasiou (1986) maturity model. However, the productive use of expertise, defined as the extent to which individuals are able to function at their highest skill level, is influenced by organization variables like role differentiation and task specialization. Role differentiation addresses the degree to which individuals are able to focus on specific jobs, with the understanding that developing and applying the greatest expertise

requires full-time effort in a given domain. Frequently, in emerging or small organizations, one will find an executive director who is responsible for multiple functions like financial management, fund development, and program implementation. Executive directors with staff support in each of those areas will be able to function differently than directors who do not have such support.

Task Specialization is used by the SOC to characterize the degree to which the organization supports individuals to work at higher order tasks within a domain. An example of task specialization familiar to most is the hierarchical specialization of tasks in the medical field in which several levels of skills are employed to free physicians for higher order tasks of diagnosis and treatment.

Infrastructure (5A) and Technology (5B) were added to SOC 11 as sub-features of *Core Services* because of the strong relationship they have to organizational capacity. The SOC 11 operational definition of infrastructure includes building or office space, equipment, and size of staff. Technology is treated as a sub-feature of *Core Services*. While technology might usually be considered as an element of infrastructure, it was thought necessary to treat it directly since many T & TA interventions address technology.

The SOC 1 feature, *Financial Resources* (number 2 on Table 4.2), initially derived from systems and organizational life cycle literature, was intended to describe the importation of resources from the environment. According to open systems theory, organizations need to derive resources from the external environment in exchange for their work in order to survive (Bertalanffy, 1968). More mature organizations could be expected to have more stable and adequate revenue streams (Harrison, 1987; Quinn & Cameron, 1983). Members of the RWJF program and evaluation staff were also concerned with the state of development of financial management systems. Do grantees or potential grantees possess the internal controls and mechanisms to ensure that grant funds are spent as intended? The initial thinking was that the answer to the practical question would be obtained as a function of the features of *Role Specialization* (number 4, SOC 1 on Table 4.2) and *Administrative Skill Development* (number 5, SOC 1 on Table 4.2). In the pilot tests, it became clear very early that the distinction between how organizations obtained resources and how they managed those resources once obtained needed to be addressed more precisely. Foundation staff as well as grantee staff needed to have a clear explanation of which issue was being addressed in these areas. We found that they frequently would be prepared to address either funding stream questions or fiscal operation questions, but not both.

Upon reflection, the organizational life cycle literature applied to internal operations like "Financial Systems" in the same way it applied to an organization's core service development, supporting the notion that not all internal operations could be expected to mature in the same way or at the same time. To distinguish fiscal operations from

financial resources, a new feature 4. *Internal Operations* was developed for SOC 11 with "4A. Financial Systems" being one of the internal operations and treated as a SOC feature (see Table 4.3). The insight from the need for this distinction led to further examination of what was more generally intended by the SOC 1 feature 5. *Administrative Skill Development.* It was determined that information about how nonprofit organizations go about developing administrative infrastructures appeared to be related to organizational maturation and would be useful in the maturity model context. Re-reviewing the organizational literature led to a greater appreciation of the crucial nature of cross-subsidization to the development of administrative infrastructure in the early stages of development of nonprofit organizations. Internal operations, like financial systems, MIS, and human resources systems, do not usually start as formal, identifiable systems. They tend to start as tasks added to an employee's existing job. When internal operations do begin to emerge as stand-alone functions, they are frequently cobbled together by combining resources from several core services to produce an internal operation that is more productive or consistent than having the functions distributed among the core services. Internal operations appear to emerge based on need and opportunity (mix of funding), accounting for the high degree of variability observed in nonprofit administrative configurations. *Internal Operations* development can be viewed just like that of Core Services, and so the SOC 11 applies the same sub-features to describe each of the five internal operations as are used for Core Services (feature 5 in Table 4.2) as shown in Table 4.3.

The five internal operations central to organizational capacity identified in SOC 11 were chosen to be consistent with the SABI training and technical assistance interventions.

Others might identify additional internal operations or use different names for their particular internal operations. The SOC 11 construction has the flexibility to

Table 4.3

SOC 11 Internal Operations and Sub-features

Internal operations	**Sub-features**
A. Financial Systems	A. Infrastructure B. Technology C. Role differentiation D. Task specialization E. Expertise 1. Domain expertise of key staff 2. Functional expertise of key staff
B. Management Information Systems	A. Infrastructure B. Technology C. Role differentiation D. Task specialization E. Expertise 1. Domain expertise of key staff 2. Functional expertise of key staff
C. Communications	A. Infrastructure B. Technology C. Role differentiation D. Task specialization E. Expertise 1. Domain expertise of key staff 2. Functional expertise of key staff
D. Human Resource Systems	A. Infrastructure B. Technology C. Role differentiation D. Task specialization E. Expertise 1. Domain expertise of key staff 2. Functional expertise of key staff
E. Development	A. Infrastructure B. Technology C. Role differentiation D. Task specialization E. Expertise 1. Domain expertise of key staff 2. Functional expertise of key staff

accommodate additional internal operations by simply identifying them and applying the same five sub-features. For example, one organization in the author's experience possessed an extensive "Planning" Department that handled functions normally identified as MIS.

Summary

The SOC 11 is different from the SOC 1 in the following ways: Uncalibrated indicators of change added predictors of direction of change from organizational development (OD) and systems literature that are not specific to any stage but but have established predictive validity for establishing direction of change that and applies to all of the SOC stages.

Financial Vulnerability was added to Financial Resources to include important newer research on nonprofit organization funding emerging from IRS form 990 studies that reports high predictive validity and has been replicated. Service System maturity dealt with structural and infrastructural development that was not limited to services. The modification more broadly accommodates research findings from OD and nonprofit studies. The same five sub-features of Core Services are also applied to each internal operation (e.g., Financial Systems, MIS, Communication, Human Resource System, Task Specialization, and Development).

4.4. SOC reliability

Concern for establishing the reliability of the SOC arose early in the development process from two sources, the SOC developer and the foundation evaluation officer. Developer

concerns focused on the validity of the SOC approach to measurement of organizational capacity. If the instrument could not be relied upon to measure the same or comparable features consistently by trained raters, its validity would be called into question. The uniqueness of the instrument required an assessment of its validity and reliability in order to interpret results and to demonstrate the degree to which the SOC might be sensitive to changes in organizational capacity, like those that the SABI project was expecting as a result of the training and technical assistance (T&TA) intervention.

The foundation evaluation officer identified the need to demonstrate the SOC's inter-rater reliability. That is, to what extent would different raters using the SOC to develop capacity profiles of the same organization produce similar profiles? This concern could influence the instrument's potential for wider use within the foundation as well as broader use by others. Both the developer and the foundation officer shared the additional evaluation-related concern that being able to report evidence of the SOC's reliability would contribute to improvement of the state of the art in assessing organizational capacity. None of the instruments used by other foundations that were reviewed in preparation for assessing SABI grantees' technical assistance "needs" reported any evidence relating to the reliability or validity of their instruments.

4.4.1. Background and assumptions for assessing SOC reliability

The very nature of the SOC and its development presented challenges for assessing its reliability. The first challenge was to develop an "operational definition" of

organizational capacity possessing sufficient precision that those using the SOC in the same context would produce similar results (Blalock, 1972). The term capacity has been vaguely and variously used for describing attributes of nonprofit organizations (see Chapter 2), and one could not expect to achieve consistency in applying the SOC without a more precise definition of organizational capacity. Therefore, the wording of the foundation's initial objective was adopted as the operational definition of organizational capacity: "The capacity to successfully complete a new grant or to expand an existing one." Building capacity, by extension, is understood to include interventions that would improve or enhance the organizational capacity of grantees to complete grant-supported projects successfully.

The second challenge in assessing the SOC's reliability was to determine whether the organizational capacity assessment results being achieved by the author were replicable. The foundation wanted to determine whether the SOC process could be transferred to other raters who could achieve comparable results or whether it was a process unique to the developer's training and experience. Unless the process could be replicated by others, the investment of resources in further development could not be justified.

Before accepting the proposed development, the foundation retained an outside evaluator to observe SOC administration and judge its potential for replicability. The evaluator, Vincent Francisco of the University of Kansas, was selected for his extensive experience working with nonprofit community agencies. As co-developer of the Community Tool

Box, an on-line capacity building resource for community agencies, Dr. Francisco possesses recognized expertise in nonprofit organizational functioning and the related issues of organizational capacity.

To study the SOC, Dr. Francisco reviewed the instrument and documentation, shadowed the developer in administering the SOC in two agencies, and interviewed the author about the instrument's grounding and development. He then reported to the developer and to the foundation his conclusion that the results of administering the SOC were, indeed, replicable (personal communication). He cautioned that the material was extremely complex and would be difficult to master for all but the highly experienced. This caution reinforced the experience of introducing the SOC to the SABI program staff as an alternative to conventional instruments. Their reluctance to embrace the SOC was due, in large part, to the instrument's complexity (the reader may refer to the discussion in Section 4.1 for more detail).

The third challenge to assessing SOC reliability was to determine what expertise would be necessary to further develop the instrument. Many reports describe assembling a panel of experts for assessing reliability of instruments in the formative stage of development. Expertise is frequently described only as possessing extensive training and experience in the applicable field (e.g. Shepard et al. [1999] define experts as those having at least 12 years' experience). However, it is possible for individuals to possess both extensive experience and training in a field and at the same time fail to become

experts. Cognitive psychology has demonstrated that the development of expertise entails more than experience and training. It involves learning to understand and solve problems in different ways. Dreyfus, Dreyfus, and Athanasiou (1986) describe a five-stage model of developing expertise, progressing from novice to expert as shown in Table 4.4 below.

Many of the distinguishing characteristics of performance at each stage in the development of expertise have been described in the research literature (Newell & Simon, 1972; Simon, 1974; Ericcson & Smith, 1991; Voss, et al., 1983; Brenner, 1984;

Table 4.4

Dreyfus Model of Developing Expertise

Stage 1 Novice	Stage 2 Advanced Beginner	Stage 3 Competent	Stage 4 Proficient	Stage 5 Expert
Simple rule-based problem solving. "Backward" thinking envisions solution and follows rules to get there.	Backward thinking with greater number of rules; emergence of situational thinking,	Situational problem solving (different rules for different circumstances), can adopt a perspective and follow it to solution	Multiple perspectives, rapid assessment of which to adopt, understanding of relevant data. Emergence of forward thinking	Intuitive forward thinking problem solving, recognizes underlying patterns

Johnson, 1988; Hoffman, 1992; Bereiter & Scardamalia, 1993; Pellegrino, Chudowsky, and Glaser, 2001; Leach, 2002). Some of those characteristics have potential implications for establishing the reliability of the SOC, among which include:

- Experts (Stage 5) solve problems differently than do those who are merely competent.
- Experts spend more time framing problems and less time working on solutions than competent performers (Stage 3).
- Experts tend to see underlying patterns and relationships, allowing them to intuitively identify meaningful information and exclude unnecessary information. The competent tend to get overwhelmed with data.
- Experts lose cognitive consciousness of the processes they use in problem solving and frequently cannot describe their own methods with specificity or accuracy.
- Expert performance is highly domain specific and does not necessarily transfer to any other domains.
- Expert performance takes a long time to develop, usually requiring 12 to 15 years of concentrated full-time effort to become a "world class" expert in a given domain.
- Expert performance must be maintained or it can be lost.

Clearly, the Dreyfus model and the lessons of cognitive psychology offer a more precise definition of an "expert" than that usually applied in conducting studies of reliability. The author concluded that the innovative nature of SOC, based on developmental staging of maturity, could benefit from a more precise definition of expertise for those administering the instrument. Several individuals with stage 4 or 5 expertise in nonprofit community organizational functioning could, reasonably, be expected to detect underlying patterns and issues in organizational capacity more precisely and effectively

than a similar number possessing equally long experience but less developed proficiency. Therefore, the author proposed using two other raters with expertise comparable to his own to refine the instrument and conduct preliminary assessments of its reliability. These were Raymond C. Logan, Ph.D., an organizational consultant with whom the developer had worked in community organizations in Pittsburgh, and Gregory H. Turner, a Research Associate in the School of Education with administrative and evaluation experience with an MBA. Both participated in several field trials with the developer over the course of two years as the SOC 1 evolved to the SOC 11.

The fourth challenge in addressing the SOC's reliability was the instrument's complexity and the dearth of systematic research on nonprofit organizational capacity and maturation. The developer decided to employ a scoring rubric approach to identifying stage-specific attributes of the SOC features. Scoring rubrics have been widely used in educational research because they become increasingly reliable and valid during their formative development (Lesgold, 2000; Frick & Semmel, 1978). A scoring rubric containing attributes thought to distinguish the stages was constructed for each of the SOC features. The attributes of each stage were initially identified from the developer's experience with nonprofit organizations and his reading of the various literatures. The attributes are intended to provide guidance in making decisions concerning stage of development of an organization for the features being profiled. Attributes are modified or added through specific experience so that the instrument evolves with use.

The fifth challenge for assessing the SOC's reliability is presented by its intended use as an

observational instrument. Observational studies have the added burden of incorporating observer agreement as an ingredient in assessing reliability of observational measures (Frick & Semmel, 1978). Reliability can be influenced by observer agreement (or disagreement) and so a consideration for assessing the SOC's reliability is to minimize observer disagreement. Therefore, a primary task in developing an observational instrument is to reduce interrater disagreement, thereby, maximizing agreement.

Summarizing the background considerations of SOC reliability

To summarize this section, development of the SOC is grounded, in part, on principles borrowed from educational research and cognitive psychology. It is an observational instrument intended to be used by raters possessing highly developed organizational expertise. The SOC employs a scoring rubric to guide raters in making decisions about the stage of maturity of organizational features they are observing. Theory leads one to the expectation that the SOC should become more reliable and valid as it evolves through use in two ways, (1) the rubric should improve both in validity and reliability as observational data are added, and (2) the observers using it will improve their consistency in scoring as they become familiar with the instrument and its use. In the remainder of this section, the rationale for using observer agreement as an indicator of reliability at this stage of SOC development is discussed, observer agreement data at three stages of SOC development are presented, and the implications of these findings for the development of SOC are discussed.

4.4.2. Observer agreement as an indicator of reliability

Demonstrating the adequate reliability of a measurement is a necessary first step for determining its validity (Schierer, 1993). Reliability has been defined variously as "freedom from random error" (Bart, 1999), "freedom from measurement error" (Cardillio & Smith, 1994), and "consistency" (Thorndike, et al., 1991) or "replicability" (Blalock, 1972). Observer agreement or consistency is a potential source of error in developing observational measures, and observer disagreements can limit a measure's reliability (Frick & Semmel, 1978). Observer agreement, as used here, is intended to measure the interrater reliability among three raters scoring the SOC. Such interrater reliability descriptions are frequently analyzed and reported as some form of interclass correlation (ICC) measure or percentage-of-agreement measure, such as Cohen's Kappa (Vogt, 1999; Huck, 2004). For this study, a simple percentage agreement is used because the data do not yet support a more stringent treatment and it is important to obtain an indicator of whether development is heading in the expected direction or not. A more rigorous assessment of interrater reliability can be supported once the SOC maturity model with its various rubrics have stabilized and raters have been highly trained. However, at this stage of development a simple percentage agreement will provide an indication of developmental progress on getting consistency across raters.

4.4.3. Data analysis

A six-by-six contingency table is employed on which two scorers' responses for a given feature are recorded (see Table Figure 4.5). The five stages of maturation of the SOC

maturity model are indicated as "Stage 1" through "Stage 5." A sixth stage, "Unrated," has been added to provide for cases in which one of the raters did not score an attribute. The *marginal agreement* for each stage is defined as the smaller of the row or column values of a stage divided by the larger of the two values (see Table 4.5). The *overall marginal agreement* is defined as the mean of the sum of the individual stage agreements (Frick & Semmel, 1978). Referring to Table 4.5, one can see on the first row that Rater 2 scored two attributes as Stage 2 that Rater 1 did not score (unrated). Similarily, on the third row, Rater 1 scored four attributes at stage 2 that Rater 2 did not score. The diagonal numbers 3, 6, and 1 represent attributes on which both raters scores were in agreement. The marginal agreements are the sums of the row and column agreements (the seventh column and seventh row). Generally speaking, finding agreement values in the range of .75 or better would indicate a high degree of agreement. Addition of the sixth stage might depress the agreement values initially, as disagreements about whether to score an attribute or not will usually be linked to the rater's familiarity with the SOC. The need for this scoring category should decrease as raters' training and experience improve (see appendix for Governance section of SOC).

Frick and Semmel (1978) place great emphasis on training raters in order to minimize disagreement when developing observational measures. Among the techniques they cite in their literature review are use of video tapes of data collection to create standard observational scenarios for use in training and criterion-related comparisons where the reference criteria are scores of an expert in the domain. I have presumed to use my scored observations as the "criterion" for this analysis.

Analyzing observer agreement on a contingency table requires ten or more scored attributes per feature in order to have confidence in the result (Frick & Semmel, 1978). Marginal agreement was selected over a related analysis, nominal agreement, because it accommodates some categorical disagreement. In other words, nominal agreement is the more stringent measure, requiring nearly perfect agreement on observations and it was expected that many disagreements on the SOC could be of only one stage without changing the underlying pattern of development.

Table 4.5

Example 6 X 6 contingency table for analyzing percentage agreement: governance for agency #3

Rater 2: Raymond Logan

Rater 1: Gregory Turner	Unrated	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	∑ Row
Unrated			2				2
Stage 1		3		1			4
Stage 2	4		6				10
Stage 3				1			1
Stage 4							
Stage 5							
∑Column	4	3	8	2			17

Marginal agreement = 1/4(2/4 + ¾ + 8/10 + 1/2) = 2.55/4 = .64

4.4.4. Data collection for assessing interrater agreement and reliabilities

Data presented were collected from four Small Agency Building Initiative (SABI) organizations in November 2000 and from two other organizations during 2001 and 2002. The four SABI organizations were observed by the SOC developer several times, and data from the files of those visits were used by the two SABI-related staff for scoring two iterations of the SOC as it was being modified. Two agencies were scored for each iteration. The developer visited two additional organizations accompanied by a different staff member, who independently completed a SOC profile (October 2002 version). In both instance both raters attended sessions and meetings together in order to reduce potential "noise" as result of observers receiving different input and as a necessity for being as unobtrusive to normal operations as possible. Several visits were made to each of these organizations for data collection and reporting.

4.4.5. Findings

The findings are presented in four tables which are discussed in order. Note that agency identification has been coded to protect anonymity. Table 4.6 reflects the percent agreement among the three raters on staging decisions for attributes of three features for two SABI organizations. Two of five features were scored (1) Governance and (2) Financial Resource Maturity. The remaining three SOC features were not scored because

the number of attributes was small, well below the ten attributes per feature as recommended by Frick and Semmel (1978) for calculating percent marginal agreement. The January 2002 iteration of the SOC was in use at the time. This is the earliest version of the SOC in which comparisons are possible owing to the rapid evolution of the SOC and training of raters to use the system.

Table 4.6

Percent agreement among raters for agencies #2 & #3

	Agency #2			**Agency #3**		
Feature	Raters			Raters		
	RS/RL	RS/GT	GT/RL	RS/RL	RS/GT	GT/RL
Governance	35%	20%	17%	63%	60%	64%
Financial Maturity	42%	69%	50%	48%	45%*	22%*

Note. SOC version of January 2002 Observations of November, 2000
*Marginal agreement underestimates agreement in this instance.

Each column in table presents the percent agreement between two raters on the features for the indicated agency. These data were derived from contingency tables as described in table 4.5. Raters RS/RL were 35% agreement in scoring Governance for *Agency #2* and 63% agreement in scoring the same feature for *Agency #3*.

The *Agency #2* was the first attempt to independently score and compare percent agreement, and *Agency #3* was the second. It is notable that the greater agreement on the

feature Governance for *Agency #3* occurred without revision of the instrument and may be attributable to: (1) practice is using the SOC, (2) differences in the organizations, (3) data in the files on this feature. Agreement on Financial Resource Maturity was consistent in scores for both agencies. However, a known limitation of the technique used for obtaining marginal agreement is that it can under-represent agreement in instances where the parties agree that no attributes for a given stage are present. This type of scoring limitation occurred only with *Agency #*3 organization. After analysis and consultation with a statistician, it was clear that the procedure was scoring an agreement as a disagreement which under-represents the actual agreement. Making a correction that reflects the agreement on the RS/GT comparison on Financial Maturity increases the score to approximately 61%, thus making the RS/GT agreements very consistent on this feature for both observations.

Table 4.7 contains the percent agreements on three SOC features, Governance, Financial Maturity, and Organizational Maturity, among the three raters for the July 2002 version of the SOC applied to two additional organizations. The revisions to the Organizational Maturity feature were a sufficient improvement that we were able to derive scores that were not out of line with those of the other features, Governance and Financial Maturity. The results of this exercise are mixed. On one hand, the RS/GT ratings were in greater agreement than on the earlier version. On the other hand, the RS/RL comparison indicates more disagreement with this version than that found on the earlier version.

Specifically, the agreement of 18% on Financial Maturity and 21% on Organizational

Table 4.7

Percent agreement among raters for agencies #4 & #5

	Agency #4			Agency #5		
Feature	Raters			Raters		
	RS/RL	RS/GT	GT/RL	RS/RL	RS/GT	GT/RL
Governance	35%	59%	62%	44%	100%	44%
Financial Maturity	18%	68%	15%	28%	70%	43%
Organizational Maturity	21%	60%	29%	35%	49%*	60%*

Note. SOC version of July, 2002, Observations of November, 2000
*Marginal agreement underestimates agreement in this instance

Maturity for *Agency #4* do not appear to indicate adequate agreement. The reader will recall that the RS ratings are those of the developer and interpreted as the criterion rating for this analysis. Therefore, it is not surprising that as the RS/GT agreements improve, the GT/RL agreements will tend to look more like the RS/RL agreements. Several factors were examined as possibly contributing to an apparent increase in the number of disagreements between the two raters. Three possible factors will be discussed, (1) the revisions might not have improved the validity of the instrument, (2) the raters could have been scoring data differently, and (3) some limitations may be due to the application of marginal agreements to this data.

1. *The revisions might not have improved the validity of the instrument.* I considered the possibility that the revisions to the SOC had an effect opposite to that intended, namely that they did not improve the validity of the instrument. However, as the percentage

agreement appeared to decrease somewhat between RS/RL, it increased between RS/GT. Also, the Organizational Maturity feature was improved to the point that it could be scored, and those scores were similar to the agreements on the other features. Therefore, reliability was unlikely to be regressing and some other factor might explain the changes.

2. *The raters could have been scoring data differently***.** If some of the organizational attributes being observed to make developmental stage decisions were ambiguous, the result might be one rater scoring one way and the other scoring another. Assume that an observed attribute could be interpreted by one rater as indicating stage two development, and the same attribute is interpreted as indicating stage three development by another. The result would appear as a disagreement. In practice, I have found instances of attributes being ambiguous when organizational features are in a state of transition. A protocol was established to guide decisions when this kind of ambiguity is encountered. The rater is instructed to determine the direction of the change, whether from a lower stage of maturity to a higher one or from a higher stage to a lower one. The attribute is scored at the beginning point of the transition (the "from" and not the "to"). Given the need to deal with ambiguity in the use of the SOC, the existence of scoring differences could not be ruled out.

3. *There are some limitations in the application of marginal agreement measures to this data.* The chief limitation of using simple percentage marginal agreement for this study is that it does not take into account the distance of disagreements. That is it does not distinguish between disagreements that are only one stage of development apart and those that are three or four stages apart. If ambiguous attributes are causing the disagreements, as described in (2) above, then one would expect to find single stage disagreements when

they occur. It would be helpful to have a measure that took into account the distance between the disagreements. Kendall's coefficient of concordance appears to be such a measure and has been used for that purpose in the medical education literature (Olson et al., 2003). However, Kendall's coefficient is for ranked data and is not appropriate for the type of rated data that the SOC produces. Therefore, a further analysis was completed on the data used for table 4.6 to assess the percentage of disagreements that were of the single-stage variety. The results of this analysis are presented in Table 4.8 below. As can be seen, single-stage disagreements account for a vast majority of the disagreements. This finding suggests to me that the SOC's interrater reliability is more adequate for the purpose of making stage of development decisions for the three features than the marginal agreement scores reflect.

Table 4.8

Percent of Disagreements That Are of One Stage

	Agency #4			**Agency #5**		
Feature	Raters			Raters		
	RS/RL	RS/GT	GT/RL	RS/RL	RS/GT	GT/RL
Governance	92%	40%	71%	71%	100%	71%
Financial Maturity	90%	100%	76%	75%	67%	50%
Organizational Maturity	100%	100%	60%	67%	75%	75%

Note. SOC version of July, 2002 Observations of November, 2000

The preceding analyses (Tables 4.6, 4.7 and 4.8) are based on ratings of site-visit data from files. In a sense, they could be considered much like the "standard patients" used in problem based learning (PBLs) in medical education and in "standard classrooms" used to train classroom observers. The data presented next consists of two observers making independent observations on-site. RS/GT scored Agency #6, and RS/RL scored Agency #1 using the October 2002 version of the SOC 11. Each SOC score was based on several site visits over time conducted to develop an organizational profile based on the SOC record.

The SOC results were shared with the agencies in presentations to staff and, in the case of *Agency #1*, the board. The reliability rating for *Agency #1* is relatively high. Cohen's Kappa of .80 or better is considered to be highly reliable for observational instruments (Frick & Semmel, 1978). The reliability rating for *Agency #6* is only moderate, with the exception of the rating of the Financial Maturity rating, which is high. Some factors that may contribute to the variation of reliability rating between the two organizations are worth mentioning. First, the two organizations vary considerably in size. *Agency #1* had an annual operating budget of $2.3 million and a staff of approximately 30 employees at the time it was observed. *Agency #6* had a budget of $11.5 million supporting a staff of more than 200 employees. *Agency* #6 was observed over a three-day period. *Agency #1* was observed for approximately the same time over a span of several weeks. *Agency #6* is multi-sited and it was not possible to visit all sites in the time available. *Agency #1* operates from a single site, and all projects were observed directly by the raters.

Table 4.9

Interrater Reliability of Two On-site Observations

	Agency #6		Agency #1	
SOC Feature	Raters RS/GT		Raters RS/RL	
	% Agreement	Cohen's κ	% Agreement	Cohen's κ
Governance	34%	0.41	89%	0.83
Financial Maturity	85%	0.84	84%	0.82
Organizational Maturity	61%	0.49	81%	0.81

Note. SOC version of October, 2002

Therefore, *Agency #*6 presented challenges to the raters in terms of its size and complexity. In analyzing the reliability ratings of this organization, the raters reflected that they had worked well as a team by alternating asking questions, giving the other some time during which to record the answer to the previous question. The effect was that they were able to cover more territory in a shorter time period than otherwise would have been possible. However, an unintended effect of this teamwork is the possibility that the two raters were attending to and recording somewhat different data. Frick and Semmel (1978) caution that observer agreement cannot be established when the observers are observing different phenomena. Inspection of the contingency table for the Governance feature revealed that all of the disagreements occurred when one rater did not stage-score an attribute while the other rater did supporting the possibility that the raters were attending to different information. This was not the case with the Organizational Maturity

feature, where all of the disagreements were of the single-stage variety discussed earlier.

Summary

SOC reliability was estimated on three successive iterations of the SOC 11, the January 2002, the July, 2002, and the October, 2002 revisions. Field notes from two SABI grantee site visits was scored by the developer and two co-workers in January and two other grantees were scored in July as part of the process of pilot testing versions of SOC 1 through SOC 11. A final set of observations was conducted in the field with the developer and a co-worker making independent observations in each case, using the October, 2002 revision of the SOC 11.

The January version of the SOC produced agreements among all raters on Financial Maturity, ranging from 42% to 69% with one exception (22%) between two raters on *Agency #*3. That score reflects a limitation in marginal agreements occurring when two raters fail to score in contiguous stages leaving a center stage blank. Agreements on Governance varied with higher and consistent agreement on *Agency #*3 and both lower and inconsistent agreement on *Agency #*2.

The agreements on the July version of the SOC on three features; Governance, Financial Maturity, and Organizational Maturity, produced variations similar to that of the January version, and did not produce acceptable reliability. However, the scores did reveal that

the developer and one rater improved reliability with agreements reaching 60% or better on most features for both agencies.

The developer and one rater (GT) independently scored the October version of the SOC for *Agency #*6 during a site visit. Interrater reliability was estimated using Cohen's Kappa. The following values were obtained: Governance .41, Financial Maturity .84, and Organizational Maturity .49. Similarly, the developer and the other rater (RL) independently scored SOC 11 at site visits to *Agency #*1, with the following results: Governance .83, Financial Maturity .82, Organizational Maturity .81. Agreements of .80 are considered acceptable agreement on observational instruments by Frick and Semmel (1978).

These quantitative results fail to establish that the three raters were able to consistently rate an organizations stage of maturation with different iterations of the SOC. This result was not unexpected and may be attributable to several factors discussed as follows:

1. *A flawed test*. The initial scores were based on field notes of the developer and not independent observations.
2. *Insufficient rater training*. The scorers might not have been adequately trained to produce consistent ratings.
3. *Incomplete development of the instrument*. The SOC is still evolving and observational development theory would suggest that it will improve in reliability with each revision.
4. *Flawed or inappropriate measure of reliability*. Percent marginal agreement and Cohen's Kappa are appropriate for rating scales like classroom observational scales that

have anchor points. They have the property of under-estimating reliability if raters both agree that one of the features has no attributes of the same stage. No score will be indicated for that stage. The effect is to increase the number of marginals (the denominator in the formula) and a marginal value of zero. A second flaw with the measures is that they do not take into account the distance of disagreements. Single stage disagreements are treated the same as disagreements of two or more stages. Table 4.7 shows that single-stage disagreements are a large majority of the disagreements occurring.

5. *The SOC was going through revisions when the interrater data were collected*.

The generally increasing agreement through time and the high agreement for RS/RL at *Agency* #1 are promising. Support for this view also comes from the likelihood that the RS/GT raters were likely attending to somewhat differing data at times as well as the high agreement on Financial Maturity.

4.5. Evidence of the construct validity of the SOC

This section presents evidence that the SOC measures nonprofit service organization's capacity to undertake new or expanded projects, a construct that it is designed to measure. Construct validity is required if a new instrument is measuring attributes that have not been operationally defined (Cronbach & Meehl, 1955). In their foundation paper on construct validation, Cronbach and Meehl (1955) counsel that construct validation must be investigated when no criterion is available that is fully valid. Further, they observe that, "Construct validity is not to be identified solely by particular

investigative procedures, but by the orientation of the investigator." They clearly describe the circumstance of the SOC for which no other instrument is available to directly measure the attribute of nonprofit organizations we term "capacity."

Instrument development frequently relies on the judgment of experts (defined as highly experienced and knowledgeable professionals) in relevant domains to render judgments of the validity of a new instrument to measure the attribute it purports to measure. Earlier in this chapter it was reported that an expert, Dr. Francisco, studied an early draft and observed its administration. His efforts constituted a first step of several in the use of experts for investigating the replicability and construct validity of the SOC. Evidence presented in this section is derived from two sets of professionals: (1) RWJ Foundation National Program Office staff, and (2) board and staff of nonprofit community organizations for which SOC profiles have been developed. These two groups are the primary stakeholders in the Robert Wood Johnson Foundation's efforts to fund projects in community organizations. Foundation staff members make important decisions to fund or not to fund projects and to continue or discontinue them. Their interest in the SOC has been to see the extent to which it aids the decision making process or confirms it. Community organization staff and boards are concerned with how to improve their performance or enhance their organizational development. The development of a new instrument is an expensive and time-consuming process. From a funder's perspective, it is also a high-risk investment that may or may not be of value. Nonetheless many RWJF staff members have been generally supportive of the SOC development from the outset. Clear evidence of that is best exemplified in the continuing support of the developmental

effort and the interest shown in the SOC for potential use by program and finance officers. More substantial evidence of the SOC's usefulness came from a trial of the SOC process undertaken by one of the foundation's national program offices, the Local Initiative Funding Partner's (LIFP). Three organizations were selected in which to pilot the SOC, with a deputy director of the LIFP office participating as an participant/observer. The organizations selected were small (operating budgets of $250,000 to $600,000). Two were newly formed organizations and one was about 30 years old. Two were community health clinics and one provided a range of services to cancer patients. The relatively small size of these organizations was of concern to LIFP because it suggested a greater likelihood that the organizations would not be able to sustain their new efforts. The SOC was administered during a site visit lasting approximately one and a half days during which interviews were held with key staff and board members, records were reviewed, and operations observed. The data collected were analyzed and a profile of each organization was developed and reviewed with the LIFP directors. A second site visit was made to share the results with the staff and board of each agency. The feedback from these two sources is anecdotally reported. In addition to data from the three LIFP grantees, data from three additional organizations in which the cycle of observation, profiling, and presentation to the staff and board was followed is highlighted.

4.5.1. LIFP deputy directors' feedback

Two of the LIFP organizations being assessed were experiencing difficulties while undergoing organizational changes, one was undergoing a change in executive leadership

and the other was experiencing changes in both board leadership and executive leadership. After the SOC was administered, both directors indicated that the SOC process obtained accurate assessments of organizational development and capacity very rapidly. They said that they were impressed that the process could get at basic organizational issues as rapidly as it did and uncover some of the unpleasantness associated with the changes through which the organizations were going. In the words of one, "It is astonishing how rapidly he can get them to take off their clothes and bare all in public." LIFP deputy directors both reported that the profiles did not result in any "surprises" or sudden new insights, although both indicated that they viewed some aspects of the organizations from different perspectives as a result of the SOC assessment. The lack of new insights can be considered an affirmation of the SOC's validity. Both directors had more extensive contact with these organizations than the author. They conducted on-site assessments prior to recommending them for participation as LIFP grantees. They also made monitoring visits to each grantee. Because of this familiarity, it would be unusual if they learned something that was both new and significant about these organizations from the SOC profiles. However, the author's impression was that a degree of what Zaltman termed "pseudo-clairvoyance" may have occurred. Pseudo-clairvoyance is the phenomenon of something becoming clearer with new data. A reaction like "I knew that" is an example of this kind of hindsight bias (Barrabba & Zaltman, 1991). The three organizations picked by LIFP were described as "problem children." They were considered to be questionable for continued funding. The author's observation was that the directors reinterpreted some of

their prior knowledge of the grantees organizations based on new perspective of the SOC. Some examples will illustrate this point.

A health center was considered to be deficient in financial management because it did not have a process of formally obtaining checks and balances on disbursements. However, in practice most transactions required two individuals to complete. From the SOC perspective, the checks and balances existed but they were informal. It was pointed out that organizational maturation tended to develop from the informal to the formal and that the health center practice was consistent with its early stage of development.

Another health center was serving only clients who could not obtain services elsewhere; these clients could pay only the most modest of fees. An LIFP deputy suggested that the health center's board might need to consider modifying their criteria for eligibility to receive services in order to generate more income to support its operation. From the SOC perspective, the organization was on an atypical developmental pathway. A normal pathway would be for growth in service provision to provide additional resources for organizational development. In this case, growth would have the opposite effect. Since growth would not produce new income, economies of scale could not realized. Growth in service population would only consume more of the limited resources. The SOC perspective leads to an alternative observation that a parallel activity of fund raising from different sources would satisfy the developmental needs of the organization without modifying eligibility criteria.

The third LIFP organization was a 60-year-old service organization that was experiencing declining revenues and having difficulty in executive leadership transition. LIFP was

concerned that the organization was not meeting its planned goals and objectives. Problems related to staff divisiveness and role conflict were identified as chief concerns. The SOC profile identified other areas contributing to the organization's inability to thrive that were not identified by the LIFP directors. The board of directors had decreased by 40% in the past two years. It was devolving at just the time it needed to evolve to higher level functioning. In addition, the organization appeared to be financially robust with reserves in equity and investments greater than two years' operational budget. The SOC provided a more balanced view of the organization by also identifying attributes suggesting financial vulnerability, limited revenue streams, inadequate overhead, negative operating margin for two years, and declining reserves. In addition, the SOC identified some of the core services as emerging (new) and not yet functioning at potential.

In each of the circumstances presented in these examples, the added perspective of the SOC profile seemed to moderate the concerns of the LIFP directors about the ability of the organizations to complete their grants. From the LIFP deputies' feedback, one concludes that the SOC rapidly produced an essentially accurate profile of important elements of organizational capacity for each agency, that dealt directly with its ablility to successfully carry out the grant

4.5.2. Board and staff feedback

All three of the LIFP organizations had vacancies in executive leadership at the time of the first site visit and were searching for new directors. This lack of executive leadership

undoubtedly contributed to the LIFP Office concerns about the stability of these grantees. Each handled the transition differently and all had hired new directors by the time of the return visit, which followed within 30 to 60 days of the initial site visit. All three boards reported their agreement with the SOC profile for their agency and all three reported learning something new about their organization. Some examples of benefits derived from the SOC assessment as indicated by the boards are briefly described below.

- One board was considering obtaining the services of a physician to increase cost reimbursement. The SOC revealed that the program lacked sufficient infrastructure to support the addition of a physician without loss of current production. In fact, the existing infrastructure would not support the addition of even one more nurse-practioner. The board response was to obtain additional space that would potentially double the number of examination rooms. They were also exploring the possibility of relying less on full-time staff and using more part-time staff in order to increase the number of hours that their clinics were open.
- Two of the boards expressed the hope that recently hired executive directors would be able to develop their services "to the next level" and improve fund raising to stabilize revenues. The SOC provided the insight that placing too many responsibilities on a single staff member was unlikely to result in those duties being fulfilled at the level expected. The problem was clarified as one of how to differentiate tasks to permit higher-level performance. The boards' responses to the insight differed. One board indicated that they were seeking additional funds for an extra position to support the executive director on a permanent basis. The other board decided to seek an additional board member who possessed fund raising expertise to guide development of an adequate

fund raising effort to better support the organization's operation.

- One agency had a history of starting programs and having them spin off into other organizations after they had become successful. The SOC identified this pattern as unusual for the organizational growth sought by the board. Typically, organizations use more comprehensive service offerings and serve larger client bases to generate the revenues and resources to fuel administrative differentiation and growth.
- Staff feedback was exemplified by a report from an agency executive director (ED) to a Deputy Director of the LIFP as forwarded to the author. The letter from the Deputy is excerpted here with personal identification omitted. "At our annual meeting in Stevenson, Washington, in October [2003], [executive director] of [LIFP grantee] asked me for an opportunity to meet with her privately. You will remember that she is the new ED that came on board the very week that you went to give the project feedback after completing the developmental instrument. She indicated that at first she thought it would be disconcerting to hear all about an organization so early on in her employment there. Instead, she said that she found it to be incredibly valuable and that it has helped her develop her 'change agenda' for the staff and services. It confirmed for her what she suspected from her interview process but it also helped her direct the energies of her staff. …[ED] was extremely complimentary about your contribution to the agency."

Three inferences emerge from these brief examples: (1) both those who fund projects and those who receive funds agree that the SOC accurately depicts important aspects of organizational stage of maturation for the features and attributes assessed, (2) while all indicated no major surprises in the SOC assessment of maturity, all seemed to benefit from the developmental perspective represented by the SOC, and (3) participation in the

SOC assessment is not necessarily a passive or benign experience, but can constitute a brief organizational development intervention itself. These inferences are further supported by evidence from three other organizations in which the SOC has been used in combination with feedback to the agency. The SOC was piloted in these organizations with the understanding that in exchange for the intrusion of the assessment they would receive a presentation of the findings. They were larger organizations than were the LIFP grantees with annual budgets between two and eleven and a half million dollars. The following points illustrate the experience with these organizations.

- Initial participants in the SOC assessment were limited to the leadership teams in these organizations. After the presentation and interpretation of the profile, each organization requested presentations to their general staff and to their boards of directors.
- Commenting on the SOC profile, one executive director said, "HUD [Department of Housing and Urban Development] is always telling me what and how they want me to change. Now I know why."
- One board chair reported that the board had been considering terminating the employment of their executive director but had changed their minds as a result of the SOC presentation. They saw that a lack of differentiation was impeding higher level functioning of the ED and were committed to fund additional administrative support positions to permit development of higher level administrative functioning.
- A senior staff member in one organization took exception to the SOC rating of his position. He was rated as possessing stage four expertise and experience but functioning at stage two. It was pointed out that each interview with him had been

interrupted so that he could attend to network or hardware problems that would normally be handled by junior level technicians. He smiled broadly and responded, "Oops, you got me."

- An executive director of a large substance abuse and mental health services agency called for an appointment to review "discrepancies with the SOC." Worried that the SOC profile was inaccurate or incomplete, the author promptly met with her. She revealed that the discrepancies concerning her were developmental differences in where she thought the agency was and where the SOC had demonstrated it to be. She had not disagreed with the SOC profile, but wanted to address developmental issues to strengthen her organization.
- One board had a history of heated meetings that were characterized as fatuous. The presentation of the SOC profile went smoothly, with the board asking many questions and engaging in lively discussion. After the meeting the board chair and executive director were commenting on the "good behavior" of the board. They attributed the good behavior to the developmental nature of the SOC which does not subscribe to typical value judgments like bigger is better.

All of the organizations in which the SOC was administered agreed that the resulting profile was an accurate representation of their agency. The agencies having agreed to participate in the pilot requested that the SOC results be extended to staff and to boards of directors. These examples illustrate the utility of the SOC in documenting the organizational maturity profiles.

4.5.3. Summary of Chapter 4

This chapter has presented four bodies of evidence that, when combined, cast a "nomological net" (Cronbach & Meehl, 1955) supporting the validity of the SOC to measure the capacity of nonprofit organizations to undertake new or expanded projects. According to Cronbach and Meehl (1955), the laws contained in a "nomological network may relate (a) observable properties or quantities to each other; or (b) theoretical constructs to observables; or (c) different theoretical constructs to one another." Together, they converge to "operationalize" the construct of organizational capacity and provide a framework for elaborating on, or enriching, our understanding of capacity. The first section compared the SOC with the SAT, which represents the current understanding of measuring capacity. That section revealed that the features of the SOC and SAT overlapped extensively, but addressed them from different perspectives. The degree of overlap provides evidence that the content validity of the SOC is consistent with accepted judgments of experts and typical instruments used in terms of the constituent features of organizational capacity

.

The second section described changes to the SOC as it has evolved through numerous filed tests and revisions. The development of the SOC was predicated on development of an observational instrument employing a maturity model and incorporating a scoring rubric to aid scoring decisions. Borrowed from educational research, this approach has the property of becoming more reliable and valid through time if development is systematic (Lesgold, 2003). The evolution of the SOC from early versions to later

versions supports an inference that the instrument is evolving as one might expect based on theory and past experience.

The third section reported evidence for the SOC's improving inter-rater reliability early in its development. Reliability is a necessary, although insufficient, condition for establishing a measure's validity. The evidence on SOC's reliability supports an inference that the SOC is becoming more reliable through time, as the theory of developing scoring rubrics predicts, but additional information about its reliability is needed as the instrument continues to mature and as more raters are trained.

The fourth section provided evidence that the SOC was viewed as reflecting organizational capacity by key stakeholders at the foundation level and community organizations leaders. They agreed that the SOC was accurately measuring capacity and their responses suggest that the SOC teaches us "more about" capacity and enriches our knowledge about the construct, two additional principles of Cronbach and Meehl's nomological net.

These four bodies of evidence combine to form an interlocking system of observables and theoretical constructs about the nature of nonprofit organizational capacity that support successful implementation of expanded or new demonstration projects. Some

difficulties arose from the use of conventional reliability and validity tools that were developed for measures of much simpler constructs.

The concluding chapter discusses some of the lessons learned from the investigation of SOC validity and the implications of these lessons for future research on nonprofit organizational capacity and capacity building.

5. CONCLUSIONS

Background and Purpose.

Nonprofit community organizations have an established role as "laboratories" or "incubators" in the implementation of demonstration programs or conduct of social experiments to inform policy making or contribute to understanding of social intervention treatment effectiveness (Iscoe & Harris, 1984; Leviton, 1994; Leviton & Schuh, 1999). Funding agencies typically have two major goals for projects in these settings. One is to make meaningful change in the host community or targeted population. The second is to identify or test interventions that developers believe may prove to be effective in other communities to address important social concerns, such as improving health care, reducing health risk, improving HIV education, reducing interpersonal violence, or reducing homelessness. Community agencies are frequently unable to implement programs exactly as planned, or respond rapidly to problems emerging after implementation begins, resulting in a mismatch between the intervention planned and the intervention delivered.

The Robert Wood Johnson Foundation (RWJF) is one of many funding organizations using nonprofit community agencies extensively as vehicles for conducting demonstrations of innovations in social service delivery in the area of health and health care. Not infrequently, agencies implementing demonstrations fail to implement them as planned or fail to complete them. The RWJF staff observed that failures in community agencies seemed related to

organizational capacity to undertake new or expanded projects. Subsequently, RWJF initiated an experiment, the Small Agency Building Initiative (SABI), to explore capacity building prior to grant making or grant expansion in hopes of improving the success rate when community organizations implement demonstration or new programs. The work reported here emanates from evaluation work on SABI to develop instrumentation appropriate for studying organizational capacity begun in 1999.

A maturity model approach formed the basis for development of the SOC instrument. A maturity model is defined as involving a set of *features* and a related set of *levels* or *stages* (Lesgold, 2000; 2003). A scoring rubric containing attributes of each feature for each stage of development is employed as a guide for assessing the stages of development of the features at the time of observation. One advantage of this approach was its observational nature. Observational instruments developed in this fashion usually evolve with use, becoming increasingly reliable and valid through iterative improvements in scoring rubrics and better training of observers (Frick & Semmel, 1978).

The purpose of this study was to investigate the validity of a maturity model for staging the organizational capacity of nonprofit community organizations that affects their ability to implement or sustain social interventions or demonstration programs. This chapter summarizes what was learned during the course of the study and concludes with a discussion of the implications of those lessons for future research. This discussion is guided by a slight paraphrase taken from Cronbach and Meehl (1955): "The problem is not to conclude that the instrument 'is valid' for measuring nonprofit organizational capacity. The task is to state as definitely as

possible the degree of validity the SOC is presumed to have."

Establishing the validity of the SOC as a measure of organizational capacity was a concern from its inception. Few of the instruments reviewed as possible candidates for use by the SABI project to measure organizational capacity contained evidence of their validity or reliability. When validity was discussed, it was related to face validity, agreement among "experts" that it would measure organizational capacity. The conventional approach to measuring organizational capacity was judged to be insufficient to meet the needs of the SABI project on several grounds, including unclear operationalization of what is meant by capacity and no clear demonstration of sensitivity to changes in an organization's capacity to successfully implement expanded or new projects. Instead, the development of the SOC, grounded on precepts of maturity modeling, was undertaken as an alternative to a more conventional approach. This departure from conventional measures concerned some RWJF stakeholders, most notably the program officers responsible for implementation of the demonstration initiative. Program officers, arguably, had the most invested in the successful implementation of the initiative, and some of them felt uncomfortable straying far from conventional measures used by others engaged in capacity-building activities. Their initial anxiety was to see that the SOC included or addressed those established elements or items considered by other foundations and experts as important to organizational capacity.

The SOC employed a five-stage maturity model of organizational development ranging from least developed or startup to most highly developed. The features and sub-features that characterize organizational capacity were obtained through interviews with RWJF staff and through a synthesis of a broad base of research literature including: organizational development,

management science, organizational life-cycles, systems theory, cognitive psychology, and studies of nonprofit organizations. An abundance of information was available for identifying features of capacity, but little research was found that delineated attributes of features of capacity that characterized one stage of maturity from another. A scoring rubric containing stage-specific attributes was developed to serve as a guide for identifying stage of development.

A critical element of this study was to investigate the validity of the features, sub-features, and attributes of the SOC as a maturity model. Estimating the validity of the SOC was based on evidence from four sources: (a) the initial development of the central features of organizational capacity and the maturity model, (b) an analysis of the revisions of the initial SOC based on field trials in SABI grantees and in nonprofit community agencies in Pittsburgh, (c) initial assessment of interrater reliability among the three SOC project staff, and (d) feedback from RWJF staff and community agency officials.

Conclusion 1. The extent to which the SOC measured important aspects of organizational capacity was investigated by comparing the SOC with the SAT. The SAT represented conventional instruments for assessing organizational capacity since it was a derived from several such instruments used by other foundations for that purpose. The SAT addressed 11 features of organizational capacity:

- *Organizational Purpose.*
- *Governance.*
- *Planning.*
- *Legal*.

- *External Environment.*
- *Programs.*
- *Staff/Communications/Decision-making.*
- *Public/Community Relations.*
- *Financial Management.*
- *Computer Systems and Hardware.*
- *Fund raising.*

The SOC identified six features of capacity:

- *Governance*.
- *Service Systems*.
- *Service or Programs Skill Development.*
- *Role Function and Specialization.*
- *Administrative Skill Development*.
- *Financial Maturity*.

The comparison revealed considerable overlap in the features of capacity accommodated by the two instruments, although the treatment of the features was different with the SAT tending to focus on formal development of policies and procedures and the SOC tending to focus on the dynamics of maturity of the features. The comparison supports the conclusion that the SOC provides adequate coverage of the content conventionally employed to measure capacity. The conclusion that the SOC comprehensively addresses the central features of capacity is further supported by the evidence provided by "expert" judges who reviewed the SOC as well as the

directors of those agencies profiled using the SOC who judged that the SOC provided an accurate and useful measure of capacity.

Conclusion 2. *Conceiving organizational capacity as a dynamic construct in which maturation occurs systematically has a profound effect on the ways it is understood and measured.* The SAT focus is largely on formality of organizational features, like by-laws, budget and planning processes, and policies. The organizational assessment produced by the SAT is static and describes *whether* an organization has or does not have a given feature. The SOC appears to clarify organizational capacity by providing a framework for understanding *how* organizations do things and, thereby, providing a basis for assessing organizational potential for increasing or decreasing capacity. David Wiley (1991) offers an explanation of why this difference occurs in the validation of evaluation measures when he suggests that we

> "... reconsider the concept of validity ... from the perspective of what is intended to be measured rather than the uses to which the measurements might be put."
>
> Because "…an exclusive focus on use embeds one in the analysis of the social and political processes which determine and influence those uses."

Conclusion 3. *The SOC instruments have demonstrated both their accuracy and utility in a variety of nonprofit organizations by identifying the stage of maturation of the different organizational features, permitting comparison of the variations in maturation among the organizational features, thereby enhancing understanding of potential or limitations in capacity to implement projects.* Not all features of a nonprofit organization mature at the same rate. For example, management functions tend to mature later than do their core services (Letts et al.,

1999). The SOC identifies the stage-differences in maturation providing insights into how they inter-relate in a particular organization, thus aiding organizational interventions to be appropriately focused and stage-appropriate.

Conclusion 4. *The maturity model allows observation and documentation of the dynamic nature of maturation of attributes within each feature.* Changes may be seen in some attributes of a feature before it has matured sufficiently to be scored as a stage change. For example, the addition of a support person to do routine tasks so that a bookkeeper has time to pay more attention to financial planning and monitoring did not change the overall stage of maturity of the internal financial system but it certainly added to the capacity of the financial system. Another organization added two positions to its administrative staff to develop the positions of Chief Operating Officer and Chief Financial Officer. Clearly this planned change has increased the administrative differentiation of the organization. Several SABI organizations increased the size of their financial and equity reserves, making them more financially robust. Capacity can be built in small increments that eventually accumulate to the point of representing a stage change. The observational nature of the SOC makes it sensitive to these kinds of attribute changes.

Conclusion 5. *The maturity model and life-cycle perspective of the SOC has led to an initial delineation of the importance of the relationship between core services and internal operations.* Attributes one expects to see in early-stage organizations include little differentiation of administrative functions from core service functions, and a high degree of cross subsidization of administrative operation. Less differentiation coupled

with more cross subsidization tend to indicate less maturity, greater differentiation and reduced cross subsidization tend to indicate greater organizational maturity.

The literature indicates that good management systems and practices differentiate well performing organizations from less well performing ones (Letts et al., 1999; Light, 1998). The SOC identifies a common set of features for determining the stage of maturation for both core services and internal operations. What distinguishes internal operations is the manner of their development. Funding for core services is usually available from external sources to support adequate implementation, while funding for internal operations is not available from external sources (Letts et al., 1999). Internal operations are frequently cobbled together from a variety of sources or they are cross-subsidized using resources earmarked for services. The SOC provides an initial framework for systematically studying this phenomenon in nonprofit organizations.

Conclusion 6. *Conventional views of establishing the validity of measurement instruments for assessing organizational capacity were not always adequate for investigating the validity of the SOC in its formative stage of development.* Multiple methods were employed to investigate the (1) content validity, (2) construct validity, and (3) inter-rater reliability of the SOC.

1. The content validity was investigated by comparing the SOC with the SAT. The SAT was a composite assembled using parts of existing instruments and represented a typical, if not exemplary, example of measures of organizational capacity for identifying T & TA capacity building needs.
2. Construct validity was investigated in two ways; (a) analysis of SOC's evolution by

comparing an earlier version of the instrument with a later version, and (b) expert opinion in the form of foundation staff and agency boards and staff views of the validity and utility of the SOC profiles.

3. Inter-rater reliability was estimated comparing scores of three raters at three different times in the evolution of the SOC.

No instruments purporting to measure capacity in other than static terms were found, eliminating correlation or criterion studies as methods for estimating the SOC validity. One technique for estimating validation is to identify expected organizational differences in capacity among a group of nonprofits and see whether the SOC reflected those differences. This technique requires some prior work in identifying groups of organizations or differences in capacity among a group of nonprofits with known differences in organizational capacity. The state of the art in studies of nonprofit organizations has not progressed beyond anecdotal and descriptive work, making the SOC an early, if not first, instrument providing a framework for comparing features of organizational capacity in a systematic manner. The dynamic and unique nature of the SOC makes conventional methods for estimating its validity problematic at its current state of development. Instead, the developer has fallen back on the logic of construct validation as set forth by Cronbach and Meehl (1955) and attempted to cast a nomological net to make clearer what contributes to the construct of nonprofit organizational capacity. A valid nomological network relates "(a) observable properties or quantities to each other; or (b) Theoretical constructs to observables; or (c) different theoretical constructs to one another." The SOC clearly has aspects of all three of these properties.

The construct theory of the SOC relates different theoretical constructs to one another being based on a synthesis of precepts from diverse sources including: maturity modeling, organizational life-cycle theory, systems theory, cognitive psychology, evaluation theory, stage-state change analysis, developmental physiology, and educational research. It also relates theoretical constructs to observables by staging the maturity of observable attributes of organizational capacity; and the nomological network provides a framework for relating observable properties (organizational features) to each other.

Conclusion 7. *The conventional techniques for estimating the inter-rater reliability of measurement instruments and observational scales were only partially appropriate for the SOC at its current state of development.* The SOC is a rating scale. Observers using the SOC rate organizational maturity on a five-point developmental scale, from least developed to most developed, on an array of attributes for each feature. These ratings can be viewed as ranked or ordinal in terms of the 1 to 5 stages of organizational maturity. However, a problem can occur when the scores of two raters are *ranked* on the 5 stages of the SOC on a series of attributes. Ranking usually has the raters ranking *n* attributes on a scale of 1 to *n.* There are techniques for adjusting for ties in the rankings, for example when a rater ranks two attributes as equal. With the SOC, ten or more attributes of each feature were ranked 1 to 5, resulting in many ties. Adjusting for the ties had the effect of producing either unusually low or high apparent inter-rater agreement depending on the number of attributes rated and the number of stages in which scores were assigned. The solution to this problem was to consider the SOC as a rating scale of nominal data. A second problem arises when one rater rates a specific attribute and the other does not rate it. A sixth column and row were added to the contingency table to accommodate

occurrences of this type. It was learned that when the raters were in agreement that no attribute was rated in the first stage the effect was to increase the denominator by one, thereby underestimating the actual agreement. Citations in the literature contain examples of researchers falling into the trap of applying nonparametric statistical tests developed for ranked data, like Kendall's coefficient of concordance, to rated data and reporting statistically significant reliability scores (e.g. Olsen et al., 2003). Two appropriate measures for rated data are Cohen's Kappa and simple percent agreement. The scores obtained with these measures are highly correlated, as can be seen in Table 4.8. There were difficulties with these measures when they were used to measure the SOC's inter-rater reliability. These problems were not described well in the nonparametric literature examined. The first problem relates to how one should handle disagreements in which one rater scores an attribute of a feature and the other rater does not score it. Examples of this type of disagreement can be seen on Table 4.1 which shows six disagreements of this type occurring out of a total of 17 observations. Table 4.1 also demonstrates how the developer decided to handle the problem. The most noticeable reductions are produced when the raters agree not to score any attribute as stage one, and yet do have some attributes in which one scores and the other does not. The reduction can be illustrated by eliminating the stage 1 scores from table 4.1. The resulting percent agreement is .45 instead of .64, the difference between encouragement and discouragement for a developer.

The second problem relates to the practical problem differentiating between disagreements that are only one stage apart and those that are two or more. One would expect more one stage disagreements than two stage disagreements, if raters are observing general patterns of organizational development. Further more, a few one stage

disagreements would not greatly change the central pattern of an organization being observed.

A system that weighted disagreements by how many stages apart they differed would have improved the assessed reliability of the SOC, as illustrated on Table 4.7. Cohen's weighted Kappa was developed for this kind of weighting, but it was developed for ordinal data. The weighted Kappa is a modification of his Kappa for nominal data. It employs quadratic weights to adjust for the degree of disagreements and might serve in future iterations of the SOC when the instrument has stabilized and will support a more stringent treatment.

Future research

The SOC is a work-in-progress and more developmental work is necessary before it is completed. The next step is to construct scoring rubrics for internal operations and for core services and refine the rubrics for governance. As work on these advances one can expect interrater reliability to continue to improve. A second immediate task is to train additional raters. It is known that the reliability of ratings increase as the number of judges or observers increase (Haynes & Zander, 1953). In addition to improving the metrics of the SOC, this researcher believes that the evidence suggests that increasing the number of organizations profiled will accelerate knowledge about nonprofit organizational maturation by expanding the nomological network cast by the SOC.

The SOC provides a more highly differentiated construct for conceptualizing nonprofit

organizational maturation than those usually described in capacity-building evaluations. This construct provides a framework that allows systematic study of the interactive effects of varying rates of maturation of different features within an organization. For example, a three-stage difference in maturity among the SOC features *core services*, *financial systems*, and *governance* has been observed in several SABI grantees. There appear to be some interactive patterns of maturity in the variations in development of these features that have not been described or explained in the nonprofit or organizational literature. For example, on several occasions one or more of a nonprofit's core services have been observed to mature from a stage one to stage three before administrative differentiation matures from stage one to stage two. Upon reflection this observation makes sense because small nonprofits tend to rely heavily on cross-subsidization to support development of administrative infrastructure. The SOC differentiated framework may lead to studies identifying varieties of asynchrony in nonprofit organizational maturation in the same way that life cycle research did in medicine and psychology. For example, in medicine, life cycles studies of asynchronous development lead to two sub-studies of development (1) within individuals and (2) among groups of individuals (Dragastin & Elder, 1975).

Finally, the use of a maturity model produced a differentiated evaluation construct that appears to improve the typical "operationalization" of "capacity" and "capacity-building" which tends to be molar (see discussion in Chapter 1). Differentiated evaluation constructs have been recognized as important for improving the likelihood for successful evaluations of complex interventions (Lipsey, 1989; Rutman, 1980; Shadish, Cook, & Campbell, 2002). Two additional areas of study are indicated from the perspective of improving methods of evaluating large molar constructs and complex interventions.

First, the SOC is still being developed and the rubrics necessary to improve observer judgments for attributes of features of *internal operations* and for *core services* need further work and validation for the instrument to emerge as a truly comprehensive instrument. Second, study is needed to determine the conditions under which less expert observers can be trained to obtain reliable scores on the SOC.

Summary

The SOC appears to be improving in its reliability and validity with each revision consistent with the expectations of developing observational instruments employing scoring rubrics. Several areas of additional research and development that could improve its reliability and validity will be summarized and briefly discussed: (1) more research on stage-specific attributes for governance and internal operations, (2) training additional raters, (3) studies of interactive effects of variable maturation of features for patterns or varieties of asynchrony, and (4) longer term follow up with agencies already profiled.

(1) *More research on stage-specific attributes for governance and internal operations*. The literature on organizational life-cycles and maturity models upon which the SOC was based suggested that nonprofits mature by stages. Experience with the SOC confirms that patterns of development can be described. However, there is virtually no body of systematic research identifying attributes of each stage and attributes of transition from one stage to another. Work in this area should improve our ability to develop rubrics for recognizing stages of development.

(2) Training additional raters. Training additional raters will provide a better estimate of inter-rater reliability according to Heyns and Zander (1953), who counsel that it is

necessary to establish the consistency of observers before one can establish the consistency or inconsistency of the behavior being observed.

(3) *Studies of interactive effects of variable maturation of features for patterns or varieties of asynchrony.* A principle of construct validation inferred from Cronbach and Meehl (1955) is that the SOC generates nomologicals or hypotheses that are confirmed by observation or experience. Experience with the SOC has produced tantalizing hints that further study of the interaction of variable degrees of maturation among features will generate hypotheses about nonprofit organization that can be studied systematically and objectively. The example above, about variable maturation of core services, financial systems, and governance could have potential for developing stage appropriate interventions if verified.

(4) *Longer term follow up with agencies already profiled*. Extending the follow up of organizations could prove highly important to the development of the SOC. The strength of the SABI T&TA intervention was relatively modest. Organizational changes can be slow in cross subsidized environments. Obtaining longer term follow up data would have value in determining whether accumulation of small changes over time produces overall stage changes.

APPENDIX A

Needs Identification and Self-Assessment Tool

DRAFT Needs Identification and Self-Assessment Tool[1]

CONFIDENTIAL

November 20, 2000

Date: _____

Organization Name: _____

Contact Information: _____

Date of Organization's Inception: _____

Executive Director: _____

Number of Employees: _____

Annual Operating Budget: _____

Number of Board Members: _____

NEEDS IDENTIFICATION AND SELF-ASSESSMENT TOOL COMPONENTS[2]

[1] This tool was adapted from the National Endowment for the Arts' Organizational Self-Assessment Tool, LISC's Capacity Building Survey, and the Irvine Foundation Assessment Tool.

[2] Please rate each item on the tool, where requested, as Well Developed, Adequately Developed or Needs Development based on your view of the current situation of your organization.

ORGANIZATIONAL NEEDS IDENTIFICATION AND SELF-ASSESSMENT CHECKLIST

ORGANIZATIONAL PURPOSE	WELL DVLPD	ADEQUATELY DVLPD	NEEDS DVLPMNT	N/A
1. Concise, written Vision/Mission Statements (please attach copy)	X			
2. Understanding of Vision/Mission by stakeholders: a. Board of Directors	X			
b. Staff	X			
c. Constituencies/ general public **There are many misperceptions about the Agency and who it serves (onl What services are provided?)**			X	
d. Prospective funders		X		
3. Vision/Mission statement reviewed regularly **Both Vision and Mission statements are printed on the back of ID cards and also on the back of agendas for each Board and Staff meeting.**	X			
4. Concise, written history of the organization	X			

5. Within the category of organizational purpose, what are the primary areas in which technical assistance is needed?

This is a strong area for the Agency.

GOVERNANCE	WELL DVLPD	ADEQUATELY DVLPD	NEEDS DVLPMNT	N/A
1. Written Board member job descriptions and understanding of responsibilities (please attach list of Board of Directors, officers and committees)		X		
2. By-laws: current, functional, used (please attach copy)	X			
3. Procedures for evaluating short-term/long-term objectives		X		
4. Procedures for recruiting and orienting new Board members		X		
6. Procedure for evaluating Board members and/or Board rotation		X		
7. Structure of Board committees		X		
8. Frequency and attendance at Board meetings		X		
9. Procedure for meeting notice and preparation		X		
10. Effective use of time at Board meetings (please attach minutes of the three most recent meetings)		X		
11. Staff input into Board decision-making	X			
12. Financial reports presented to and understood by Board	X			
13. Method of reviewing auditor's report	X			
13. Board represents community served **The Board covers four counties /** ; **The Executive Committee is seeking to include more minorities and women.**			X	
15. Board's understanding of Board/staff relationships		X		
16. Board's understanding of fundraising **Only in the last year have Board members begun being requested to make contributions.**			X	
17. Board members make sufficient cash donations annually **New request since last year for members to contribute to the Agency.**			X	

Governance continued

18. Board members give time/expertise		X		
19. Board members advocate for the organization in the community	X			

GOVERNANCE Continued	WELL DVLPD	ADEQUATELY DVLPD	NEEDS DVLPMNT	N/A
20. Board members review/approve long range plan	X			
21. Board members recruitment of new donors		X		
22. Board members sense of involvement		X		
23. Board members sense of accomplishment		X		
24. Role of advisory committees				X
25. Within the category of governance, what are the primary areas in which technical assistance is needed? **The Agency needs to continue to develop strategies to diversify Board and get greater involvement with fundraising.**				

PLANNING	WELL DVLPD	ADEQUATELY DVLPD	NEEDS DVLPMNT	N/A
1. Comprehensive, multi-year organizational plan in place and in use		X		
2. Plan reviewed and updated annually with Board and staff		X		
3. Evaluation of previous year activities in relation to plan **The Agency needs help with getting the best type of data collection for services. Much of the work is currently done manually.**			X	
4. Within the category of planning, what are the primary areas in which technical assistance is needed? **The Agency needs tools to conduct evaluation and demonstrate effectiveness.**				

PROGRAMS	WELL DVLPD	ADEQUATELY DVLPD	NEEDS DVLPMNT	N/A
1. Sense of vision and continuity provided by organization's leadership		X		
2. Annual program planning process	X			
3. Written annual program plan	X			
4. Formal communication of annual program plans to staff / volunteers		X		
5. Formal communication of annual program plans to Board		X		
6. Method for review/evaluation of prior year's program		X		
7. Method for determining whether programs meet constituent needs **The 'needs development' is based on needing a simplified form to get those responses. [illegible] leadership is very involved in the planning process community-wide.**			X	
8. Written multi-year program plan		X		
9. System for costing-out program elements		X		
10. Volunteer support of programs			X	

11. Within the category of programs, what are the primary areas in which technical assistance is needed?

Technical assistance for coordination of volunteers, training and development (there are issues with confidential nature of the l[illegible] work).

STAFF/COMMUNICATIONS/DECISION MAKING	WELL DVLPD	ADEQUATELY DVLPD	NEEDS DVLPMNT	N/A
1. Clarity of reporting relationships (please attach organizational chart / hand-drawn is acceptable)		X		
2. Current and accurate written job descriptions (please attach job descriptions for leadership and management positions) **These are in the process of being changed to performance based.**		X		
3. Administrative leadership	X			
4. Communication among staff members				
5. Communication between staff and Board **Agency-wide needs development; within excellent.**	X			
6. Staff's understanding of the function and duties of the Board		X		
7. Staff participation in planning		X		
8. Appropriateness of staff size re: programming challenge **Funding is a major problem, particularly for professional staff.**			X	
9. Staff experience in relationship to the job assignment		X		
10. Upgrades from volunteer to staff				X
11. System for volunteer recruitment (i.e. volunteer manual)			X	
12. Description of roles and responsibilities of volunteers			X	
13. Time available to perform jobs satisfactorily		X		
14. Opportunities for staff training and personal development		X		
15. Systems for recruiting/using/thanking staff & volunteers			X	
16. Written personnel manual and policies **The salaries are a major problem. They are too low, but they have excellent benefit packages.**	X			

17. System for annual performance reviews **A new system is under development which will simplify performance reviews.**		X		
18. Benefit package for staff	X			

STAFF/COMMUNICATIONS/DECISION MAKING Continued	WELL DVLPD	ADEQUATELY DVLPD	NEEDS DVLPMNT	N/A
19. Salary scale for staff **The salary scale is a problem for staff retention. Staff cannot afford to stay due to low salaries.**			X	
20. Staff use of consultants/contractors		X		
21. Staff morale		X		
22. Office equipment		X		

23. Within the category of staff/communications/decision making, what are the primary areas in which technical assistance is needed?

Recruitment and adequate utilization of volunteers.

PUBLIC / COMMUNITY RELATIONS	WELL DVLPD	ADEQUATELY DVLPD	NEEDS DVLPMNT	N/A
1. Consistency and clarity of organizational image communicated to public			X	
2. Organizational chart clearly depicts structure of agency	X			
3. Mechanism for handling of public queries, complaints, etc.		X		
4. Annual public relations plan			X	
5. Effectiveness in representing the organization: a. to its constituencies			X	
b. to the general public			X	
c. to the press			X	
d. to critics			X	
6. Mechanisms for receiving and evaluating feedback from various sources **It is stronger from clients and specific funding sources than from general public.**		X		
7. Organization's publications			X	
8. Size of staff in relation to task			X	
9. Experience of staff in relation to job assignment			X	
10. Budget for public relations **There is neither budget nor staff for PR. Responsibilities are carried by Assistant Executive Director.**			X	

Public / Community Relations continued

11. Within the category of public / community relations, what are the primary areas in which technical assistance is needed?

The Agency is just beginning to develop a PR function as part of responsibilities of the Assistant Executive Director who is a [redacted] Programs. Therefore assistance here would be appreciated. They need help better telling the story of [redacted] This would help with funding

[redacted]entioned Agency that developed an "accountability report" which helped attract funds to agencies.

FUNDRAISING	WELL DVLPD	ADEQUATELY DVLPD	NEEDS DVLPMNT	N/A
1. Process for setting annual fundraising goals , **'Needs development' is considering new funds or funds not specified by Government for specific programs.**			X	
2. Mechanism for providing development staff input into fundraising goals **This is a new responsibility of the Assistant Director and needs help with this.**			X	
3. Mechanism for providing Board input into fundraising goals			X	
4. Ability of organization to consistently meet fundraising goals **There are no fundraising goals set yet.**				X
5. Written development plan for annual operating support **There is no plan for development yet.**				X
6. Plan indicates secured, renewable and speculative contributed income			X	
7. Size of development staff in relation to task			X	
8. Experience of staff in relation to job assignment			X	
9. Grant-writing expertise	X			
10. Individual donor research capacity			X	
11. Corporate, foundation, government grants research capacity			X	
12. Involvement of Board in prospect identification			X	
13. Involvement of Board in individual donor solicitation			X	
14. Non-Board volunteer support of fundraising effort			X	
15. Solicitor training			X	
16. Earned income/contributed income mix			X	

Fundraising continued

17. Ability of organization to set and accurately project earned income goals			X	
18. Ability of organization to consistently meet earned income goals [where applicable] **Goals are not yet established.**				X

FUNDRAISING Continued	WELL DVLPD	ADEQUATELY DVLPD	NEEDS DVLPMNT	N/A
19. Organization's ability to attract: a. individual donations, gifts			X	
b. corporate and corporate foundation gifts			X	
c. private foundation funding			X	
20. Level of local government grants			X	
21. Level of state government funding			X	
22. Level of national government funding			X	
23. Development materials			X	
24. Computer support for fundraising			X	
25. Computerized procedures for updating/purging information			X	
26. Gift recording system			X	
27. Gift acknowledgment system			X	
28. Systematized information about donors/prospects			X	
29. Mechanism for systematic renewal of past/lapsed donors			X	
30. Systematic plan for expanding donor base			X	
31. Budget for fundraising expenses			X	

Fundraising continued

32. Within the category of fundraising, what are the primary areas in which technical assistance is needed?

The [redacted] spends ¼ to 1/3 of [redacted] time to fundraising tasks. They have good grant-writing skills. Funding diversification would be a strategic goal. There is no development plan or budget. Technical assistance to do this is a priority for [redacted] particularly since State grant will end in one month.

FINANCIAL MANAGEMENT	WELL DVLPD	ADEQUATELY DVLPD	NEEDS DVLPMNT	N/A
1. Annual budgeting process (please attach a copy of your annual budget and latest budget-to-actual report)		X		
2. Computerized accounting/budgeting/reporting systems		X		
3. Board mechanism for reviewing monthly statements		X		
4. System for regular cash flow projections and monitoring		X		
5. Ability of organization to manage cash flow		X		
6. Formalized cost controls		X		
7. Ability of organization to pay accounts payable within 30 days		X		
8. Annual financial audit (with management letter) by CPA (please attach most recent audited financial statements)		X		
9. Board system for reviewing audit		X		
10. System of internal controls		X		
11. Payroll tax deposits made when due		X		
12. Ability of organization to manage debt repayment of notes/loans outstanding		X		
13. Line of credit for the organization		X		
14. Investment and endowment returns **Currently, none exists.**				X
15. Formal policy that maintains integrity of • cash reserve • endowment		X	X	
16. Size of staff in relation to tasks			X	
17. Experience of staff in relation to job assignments		X		
18. Staff's understanding of financial systems		X		

FINANCIAL MANAGEMENT Continued	WELL DVLPD	ADEQUATELY DVLPD	NEEDS DVLPMNT	N/A
19. Within the category of financial management, what are the primary areas in which technical assistance is needed? **The Agency needs help with development non-restricted funds and help with figuring out what an appropriate development budget should be for this Agency, with appropriate staffing.**				

COMPUTER SYSTEMS AND OPERATIONS[3]	WELL DVLPD	ADEQUATELY DVLPD	NEEDS DVLPMNT	N/A
1. Computer systems / operations supports organization's: a. planning		X		
b. programs			X	
c. communications		X		
d. staff / volunteers / Board members			X	
e. financial management		X		
2. Computer training for staff / volunteers / Board members **Some staff members do not know how to use existing system. This is a problem of time/staff needs.**			X	
3. Computer software systems **has a new software program that has a practice management approach but all staff have not been trained in it yet.**			X	
4. Computer hardware acquisitions **Some hardware is outdated. They just received a small grant to help upgrade some of their hardware.**			X	
5. Within the category of computer systems and operations, what are the primary areas in which technical assistance is needed? **The Agency needs technical assistance for additional hardware and more opportunities for staff training (they would also need volunteers or funding for temporary staff to free up full-time staff to attend training). The Agency needs training in billing processes; the software already exist (office guide by Princenet at about $10,000 per site with $2,000 maintenance contact – usually about $50,000). Their hardware needs upgrading to work with practice management software as well as hardware used for fund development.**				

[3] Please note, you will be e-mailed an on-line survey that will allow for an in depth understanding of your organization's computer systems and operations.

LEGAL	WELL DVLPD	ADEQUATELY DVLPD	NEEDS DVLPMNT	N/A
1. Percentage of budget allocated to insurance policies		X		
2. Access to legal counsel and expertise (Board Member, paid or pro bono services)		X		
3. Staff participation in training on compliance issues **A compliance officer will be on-staff full time beginning in January and will train staff in compliance issues.**			X	
4. Existence of personnel and grievance policies/procedures		X		
5. Existence of appropriate insurance policies		X		
6. Compliance with reporting other legal requirements		X		
7. Achievement of identified legal goals		X		
8. Does the organization have any outstanding lawsuits (yes or no)' **has no outstanding lawsuits. There are two outstanding for**				

EXTERNAL ENVIRONMENT	WELL DVLPD	ADEQUATELY DVLPD	NEEDS DVLPMNT	N/A
1. Understanding of local/regional: health, social service and economic climate	X			
2. Organization's knowledge of/relationship to immediate community	X			
3. Organization's knowledge of/relationship to municipality and county		X		

FINAL COMMENTS
1. Please prioritize the three most pressing capacity building needs identified by the staff: a. **The Agency needs technical assistance for development strategies and diversification of funding base and Board diversity.** **b. The Agency needs a human resources consultant specializing in community based health clinic staffing. Technical assistance is needed for staff retention, salary issues and finding ways to bring in additional professional component and related funding issues with supporting technology.** **c. The Agency needs technical assistance for public relations / education of community at large with view to increase funding.**
2a. Has your organization received technical assistance and/or training in the past? **The Agency has had workshops with RWJF and data collection information within context of a conference. Otherwise they have not had on-site technical assistance (vast majority of funding goes into salaries).** 2b. If so, what was purpose and who was the provider of this assistance and/or training?
3. Please comment on any additional areas of concern that are not covered by this survey.
LIST PARTICIPANTS / POSITION IN SELF ASSESSMENT ON (date 11/30)
1.
2.
3.
4.
5.
6.
How long did it take (in person hours) to complete this survey?

APPENDIX B

Staging Organizational Capacity (SOC 1)

Second Draft Grantee Assessment Instrument
November 2, 2000

Introduction

The Small Agency Building Initiative (SABI) of the Robert Wood Johnson Foundation is developing this instrument to measure developmental outcomes resulting from capacity-building projects. You are asked to rate your agency on five features and 12 sub features. These features represent stages of development of capacity, progressing from startup or initial condition to the most completely developed condition. There will be five stages of development for each feature or subfeature. The following example provides general guidance about how to score features. The example is based on skill development and acquisition and the bullelts .

☐ Stage 1. Beginning or novice level skill indicators:

- Little previous experience, an initial or startup condition.
- Performs tasks by "following the rules" or established procedures.
- Performance is context free, the known rules are applied in all situations.

Example. New drivers are instructed to maintain a one-car length interval between their car and the one in front of them for every 10-mph of speed.

Stage 2. Advanced Beginning skills indicators:

☐
- Greater experience than stage 1
- Still context free decision-making but more rules of greater complexity than stage 1.
- Begins to recognize situational elements of problems.
- Backward problem solving, that is works backwards from desired goal.
- Judges performance by how well rules or procedures are followed.

Example: Knows to increase distance between cars when it is raining.

Note: An indicator of backward problem solving can be that problems are stated in terms of the perceived solution like getting a new computer as the solution to developing financial management capacity.

☐ Stage 3. Technically competent (journeyman level) skills indicators:

- Begins to understand complexity of problems as the number of elements recognized becomes overwhelming.
- Judgment is limited as accurate sense of what is important has not yet fully developed.

- Employs hierarchical procedures for decision-making (organize problem, make plan to solve it, examine most important factors that fit plan).
- Problem solving takes the form of goals and decisions.
- Begins to recognize patterns but does not recognize full range well.

Example: A person who, in providing driving instructions, gives so much accurate but unessential detail that it confuses those attempting to follow them.

Note: Individuals at this level of skill acquisition may have difficulty establishing priorities when the problem is complex. They may spend equal time on less relevant tasks as they do on more relevant ones. Decision-making is calculated based on experience and the perspective chosen when defining the problem.

☐ Stage 4. Technically proficient skills indicators:

- Not overwhelmed with complexity because of deep involvement with task that has enabled development of a specific perspective from which salient features of problems will stand out based on this perspective. Other features will be ignored or fade into the background.
- Considerable knowledge and experience.
- Fluid performance instead of detached deliberation of choices.
- Solid pattern recognition.
- Works forward toward solutions

Example: Drivers at this level of skill development become good defensive drivers because they are able to automatically recognize potential problems and take corrective measures.

Note: Individuals at this stage of development make decisions based on both calculation and intuition.

☐ Stage 5. The technical expert or master.

- Deeply involved with vast experience.
- Does not see problems in a detached way.
- Superior knowledge.
- Remembers patterns better than Stage 4 and acquires new information more rapidly.
- Does not make decisions or devise plans in the deliberative sense, just does what normally works.
- Intuitively knows what to do based on mature and practiced understanding.

Examples: Racecar drivers report that they "become one with the car" and are unable to describe decision points. Chess masters do not think many moves ahead, they simply "play the position".

Note: True experts may not be able to describe their thought and decision-making processes as they "just do it".

These general descriptions will be applied in the rating instrument that follows.

SABI Capacity Assessment Instrument
Draft of October 31, 2000

This instrument seeks to rate organizational level of development based on the knowledge and observations of the raters. Our goal is to develop an instrument that will produce comparable results when used by different observers. You are asked to identify the developmental level of the organization being rated and provide an indication of why you have arrived at that score.

For each level one or more characteristics typical of a board at that level of development is provided as a guide for scoring. As this instrument is in development a comprehensive list of characteristics is not intended. It is also likely that and organization may possess some characteristics of more than one level. Please place a check mark indicating all that you think apply. At the bottom of each feature please rate the level that best represents the current level of development of the organization. Space is also provided for you to enter characteristics that influence your rating but are not included on the list.

Check any and all features that you observe to describe the agency even if they represent different levels or stages. You will also be asked to provide an overall rating for each feature in the space provided.

Person completing this instrument

Rater: _____

Agency: _____

Date: _____

Feature 1. Governance. This feature relates to the organization's board of directors and its capacity to conduct business.

Level 1. Initial or early stage organization.
____ Startup or young agency
____ Unincorporated or recently incorporated
____ Initial Board, not elected but selected to establish organization
____ Meetings held infrequently or on an ad hoc basis
____ Democratic decision-making process with no quorum requirement

Level 2. Advanced young or small agency.
____ Incorporated (please indicate year _____)
____ Board comprised of members recruited by the Executive Director or members
____ Meetings held at irregular intervals as needed
____ Monitoring is informal

____ Board has functional officers with fixed terms
____ Policymaking limited to approval of staff recommendations

Level 3. Established agency.
____ Board meets at regular intervals at least six times per year
____ Board members monitor project progress
____ Board selects Executive Director and approves senior staff
____ Board receives regular financial reports
____ Public face of agency is the Executive Director

Level 4. Involved board.
____ Board serves fixed terms with replacement
____ Board recruits members based, in part, on needs of organization
____ Board hires Executive Director and approves senior staff
____ Board committee structure active and functional

Level 5. Highly invested board.
____ Board active in development and promotion of organization
____ Board has dedicated staff and support
____ Board Chair active recognizable public representative (face) of organization

Rating: (circle one) 1 2 3 4 5

Please indicate factors contributing to your ranking not listed above.

Feature 2. Financial Maturity. This feature relates to the organization's ability to sustain or finance itself.

Level 1. Funding is unstable, unpredictable, or insecure
____ Startup or seed money funding only
____ Opportunistic funding (like a new or emerging policy concern)
____ Funding periods are of short duration
____ Heavy reliance on volunteer or contributed resources

Level 2. Funding limited but more stable

____ Contingency funding for activities or services but not long term (e.g. summer youth activities).
____ Projects or activities funded on a one-time only basis
____ Some support as a result of staff contribution of unpaid staff time
____ Some funded projects may not be closely fitted to organizational mission

Level 3. Established agency funding where agency has some control of own destiny.
____ Funding derived from services related to an on-going policy priority like violence prevention, or early childhood development
____ Funding for specified periods but includes provision for renewal or continuation
____ Programs funded on a fee for service basis
____ Research or demonstration projects or programs

Level 4. Stable sources of external funding
____ Agency funded through on-going source like United Way or local funding authority (Mental Health, housing, or juvenile justice)
____ Agency receives some funding from investments, endowment, or capital assets

Level 5. Highly stable funding
____ Institutionalized funding for at least a core of services or programs (e.g. funding is a line item on a state or local budget)
____ Flexible or continuing funding supported by unrelated program source like an endowment, interest income, or major fund raising activity).
____ Major source of income from sources like cost reimbursable contracts or fee for service arrangements.

Rating: (circle one) 1 2 3 4 5

Please indicate factors contributing to your ranking not listed above.

Feature 3. Service System Maturity. This feature relates to the manner in which the agency organizes their delivery of services.

Level 1. Initial or early stage organization.
____ A single project or activity

____ Two or more unrelated projects or activities (e.g. After school homework help project and a job readiness training program).
____ Activities only – no services provided (e.g. neighborhood watch)

Level 2. Services organized by project.
____ Agency has one or more projects comprised of one or more activities funded from a common source

Level 3. Agency organized into programs
____ Agency has a program comprised of two or more related projects
____ Agency receives funding from more than one source to provide the same service or set of services to different target populations.

Level 4. Agency organized into services
____ Agency provides one or more services comprised of two or more related programs

Level 5. Agency operates a service system.
____ Agency operates a system of services comprised of two or more unrelated or complimentary services

Rating: (circle one) 1 2 3 4 5

Please indicate factors contributing to your ranking not listed above.

Feature 4. Role specialization and function. This feature relates to the organization's development of administrative capacity.

Level 1. Initial or early stage organization, no administrative or service provision role specialization.

____ Administrative roles are carried out by volunteers or by staff whose primary function is to provide services.

____ Administrative functions carried out on an ad hoc basis when absolutely necessary

____ Senior staff has both administrative and service roles that often compete for the limited time of the employee

____ One or two individuals "wear many hats"

____ Staff may report that administrative tasks keep them from doing their jobs

____ Staff assigned to new projects are primarily existing staff whose primary support may well come from the new project

____ Staff assigned to new projects may have participated in proposal development while not in agency employ with the understanding that they would be employed by agency if project was funded

____ New projects suffer because assigned staff have dual roles of service provision and essential administrative tasks.

____ Staff assigned to new projects are usually paid at a lower rate than is typical for existing agencies in the area

____ Agency employs less experienced personnel for providing services

Level 2. Advanced young or small agency, administrative and service role differentiation emerges.

____ Administrative and service functions split between two individuals on a part time basis. The Executive Director will spend most time on administrative tasks

____ Administrative work is not routinized and output is irregular and uneven

____ Senior staff does not have support staff to whom tasks may be delegated

_____ Budgets do not provide overhead support for administrative positions and support

____ Project start-up is usually not on schedule because the agency must recruit new staff

____ Staff turnover is high because of low wage, atypical working hours, and extended duties not paid for from either overhead or service project budget

Level 3. Established agency, financial and administrative role differentiation emerges.

____ Agency has an Executive Director and Financial officer at least part time

____ Some support staff to carry out routine tasks and reports

____ Regular reports are completed for board, planning, and reporting requirements

____ Some administrative costs are recovered through budget process but overhead may not be a fixed rate

____ Some staff are paid from both administrative and service budgets

____ New project startup involves assignment of existing staff member to new project while recruiting replacement for that member for the former assignment

____ New project startup is likely to be on schedule but at a slower than planned pace as new line staff are sought

Lo[illegible] of senior staff member may result in significant shift in agency operations or mission.

Level 4. Service and administrative proficiency

____ ED and financial officer have support staff dedicated for routine tasks

____ Service staff supports improves productivity of professionals for things like client flow, reporting, and follow-up

____ Administrative role differentiation increased may identify staff for planning, development, communications, and/or proposal writing

____ Overhead costs are well defined and staff serving in both administrative and service capacities are billed appropriately

____ New project startup typically on time with most staff being transferred from existing assignments within the agency

____ Recruitment of staff from outside sources is easier as a result of reputation of agency

____ Staff is paid at or close to going rate for their experience and training

____ Staff turnover is low and vacancies do not cause major disruption to services of fluid delivery of those services

____ Agency is able to replace a senior staff member without major disruption

____ Staff represent a range of skills, training, and certification

Level 5. Highly productive agencies

____ Staff paid at or above average for area

____ Staff highly experienced and highly trained

____ Staff differentiation and support adequate to maximize productivity of senior staff

____ administrative oversight monitoring for all levels of performance

____ Very stable staffing

Rating: (circle one) 1 2 3 4 5

Please indicate factors contributing to your ranking not listed above.

Feature 5, Administrative skill Development

Level 1. Beginning or entry level administrative skills:
____ Little previous experience, an initial or startup condition.
____ Performs tasks by "following the rules" or established procedures.
____ Performance is context free, the known rules are applied in all situations.

Level 2. Advanced Beginning skills indicators:
____ Greater experience than level 1
____ Still context free decision-making but more rules of greater complexity than Level 1.
____ Begins to recognize situational elements of problems.
____ Backward problem solving, that is works backwards from desired goal.
____ Judges performance by how well rules or procedures are followed.
____ Problems are stated in terms of the perceived solution

Level 3. Technically competent (journeyman level) skills indicators:
____ Begins to understand complexity of problems as the number of elements recognized becomes overwhelming.
____ Judgment is limited as accurate sense of what is important has not yet fully developed.
____ Employs hierarchial procedures for decision-making (organize problem, make plan to solve it, examine most important factors that fit plan).
____ Problem solving takes the form of goals and decisions.
____ Begins to recognize patterns but does not recognize full range well.
____ May have difficulty establishing priorities because of problem complexity.
____ May spend equal time on less relevant tasks as they do on more relevant ones.
____ Decision-making is calculated based on experience and the perspective chosen when defining the problem.

Level 4. Technically proficient skills indicators:

____ Not overwhelmed with complexity because of deep involvement with task that has enabled development of a specific perspective from which salient features of problems will stand out based on this perspective. Other features will be ignored or fade into the background.
____ Considerable knowledge and experience.
____ Fluid performance instead of detached deliberation of choices.
____ Solid pattern recognition.
____ Works forward toward solutions

Note: Individuals at this stage of development make decisions based on both calculation and intuition.

Stage 5. The technical expert or master.

____ Deeply involved with vast experience.
____ Does not see problems in a detached way.
____ Superior knowledge.
____ Remembers patterns better than Level 4 and acquires new information more rapidly.
____ Does not make decisions or devise plans in the deliberative sense, just does what normally works.
____ Intuitively knows what to do based on mature and practiced understanding.

Note: True experts may not be able to describe their thought and decision-making processes as they "just do it".

Rating: (circle one) 1 2 3 4 5

Please indicate factors contributing to your ranking not listed above.

Feature 6, Service or program skill development

Level 1. Beginning or entry level service skills:
____ Little previous experience, an initial or startup condition.
____ Performs tasks by "following the rules" or established procedures.
____ Performance is context free, the known rules are applied in all situations.

Level 2. Advanced Beginning skills indicators:
____ Greater experience than level 1
____ Still context free decision-making but more rules of greater complexity than Level 1.
____ Begins to recognize situational elements of problems.
____ Backward problem solving, that is works backwards from desired goal.
____ Judges performance by how well rules or procedures are followed.

____ Problems are stated in terms of the perceived solution

Level 3. Technically competent (journeyman level) skills indicators:
____ Begins to understand complexity of problems as the number of elements recognized becomes overwhelming.
____ Judgment is limited as accurate sense of what is important has not yet fully developed.
____ Employs hierarchical procedures for decision-making (organize problem, make plan to solve it, examine most important factors that fit plan).
____ Problem solving takes the form of goals and decisions.
____ Begins to recognize patterns but does not recognize full range well.
____ May have difficulty establishing priorities because of problem complexity.
____ May spend equal time on less relevant tasks as they do on more relevant ones.
____ Decision-making is calculated based on experience and the perspective chosen when defining the problem.

Level 4. Technically proficient skills indicators:

____ Not overwhelmed with complexity because of deep involvement with task that has enabled development of a specific perspective from which salient features of problems will stand out based on this perspective. Other features will be ignored or fade into the background.
____ Considerable knowledge and experience.
____ Fluid performance instead of detached deliberation of choices.
____ Solid pattern recognition.
____ Works forward toward solutions

Note: Individuals at this stage of development make decisions based on both calculation and intuition.

Stage 5. The technical expert or master.

____ Deeply involved with vast experience.
____ Does not see problems in a detached way.
____ Superior knowledge.
____ Remembers patterns better than Level 4 and acquires new information more rapidly.
____ Does not make decisions or devise plans in the deliberative sense, just does what normally works.
____ Intuitively knows what to do based on mature and practiced understanding.

Note: True experts may not be able to describe their thought and decision-making processes as they "just do it".

Rating: (circle one) 1 2 3 4 5

Please indicate factors contributing to your ranking not listed above.

APPENDIX C

Staging Organizational Capacity

Version 11 (SOC 11)

Features		Agency Stage of Development Revision 12.1 April 27, 2004				
		Stage 1 Initial, start-up or small agency	Stage 2 Advanced young or small agency	Stage 3 Established agency	Stage 4 Proficient agency	Stage 5 Highly productive agency
1. Governance						
Environmental relationships	A	Unincorporated, recently incorporated, or start up agency.	Incorporated, Bylaws informally followed. Licenses or certificates adequate for requirements of services.	Incorporated with functional by-laws - Licenses or certificates exceed minimum for services. Provision for monitoring quality	Incorporated, working by-laws, fully certified, with a well developed set of policies and procedures in place and working.	Exemplary organization that may have an empowered executive board functioning between meetings of the full board.
	B	May be sponsored by an established agency or may be a component of a parent organization.			May sponsor start-up organization or community agency under umbrella of agency.	
	C	Public face of the organization is its mission or interest in a "break through" service or methodology, either in terms of the problem, service approach, or policy priority.	Public face of agency is the founder or Executive Director.	Public face of organization is agency name and service delivery system.	Public face of agency expands to include the senior staff.	Public face of agency will be the chair and members of the board as well as highly respected senior staff.
Board of Directors	D	May have an advisory board or small board of convenience - usually selected by founder.	New board members recruited by Executive Director or existing members.	Expanded board recruited under provisions of agency policy, may include a nomination and election process.	Members recruited, in part, on needs of organization. Members may be recruited in advance and prepared for board service through committee work.	Board members recruited based on their connections in other organizations or activities relevant to agency mission (networking).
	E	Board selected because of willingness to support efforts of founder or proposed mission.	Board selected because of support for establishing agency or support of proposed mission.	New board members may be selected for the experience and expertise they bring to the agency (e.g. attorneys or accountants).	Board members selected based on their positions as much as their experience and expertise (informal linkages).	Board members may represent formal linkages with other organizations (e.g. professional associations) or bring credibility to agency mission.
	F	Board or advisory members primary interest is in furthering organization mission.	Board members may be better advocates for mission than for agency in community. ======>	Board concerned with developing strong internal policies and procedures.	Board equally concerned with internal and external policy issues.	Board involved more with strategic policy involving external or environmental relations than with internal policies & procedures which are delegated to senior staff.

Features		Agency Stage of Development Revision 12.1 April 27, 2004				
		Stage 1 Initial, start-up or small agency	Stage 2 Advanced young or small agency	Stage 3 Established agency	Stage 4 Proficient agency	Stage 5 Highly productive agency
	G	Individuals have little experience in forming new organizations; their primary experience is with the target population, area, or service.	Board member experience is essentially as staff in domain-related services or long term involvement with agency as it has developed.	Board expansion includes members with administrative level experience or board level-experience.	Board includes some members highly experienced in both service and administrative domains.	Board members highly experienced in strategic development and forecasting. Board highly involved in ensuring future of agency and promotion of mission.
Internal functioning and organization						
Meetings	H	Meetings held on an ad hoc basis - contingency agenda usually developed by founder or Executive director.	Meetings held at regular intervals - agenda developed but loosely followed or frequent departures may cause meetings to be lengthy.	Regular meetings with agenda provided in advance and closely followed.	Regular meetings with agenda consisting of committee reports.	Meeting interval may increase to quarterly or semi-annually to review work of staff.
	I	Board meetings can be informal and may consist only of the founder polling members individually.	Board meetings may not manage time or members well. May be habitually protracted.	Meetings may have an agreed upon time limitation.	Meeting time is reduced as most debate takes place at committee level. Considerable staff support in providing background material, notices, and drafts of minutes and agenda.	Board meetings may be in form of dedicated retreats or dinner meetings.
Bylaws	J	No by laws.	Bylaws existbut informally followed	Formal Bylaws that are followed.	Formal bylaws supplemented by formal policies and procedures.	Bylaws primarily ceremonial as they are seldom necessary.
Minutes & Reports	K	Minutes not kept. May be produced to meet funding or legal requirements.	Minutes kept informally.	Board keeps a journal of minutes.	Board journal and annual report available for inspection.	Minutes, annual reports, and monographs kept. In addition, board actions may be released to media as newsworthy policy statements.
	L	Reporting to board informal, ad hoc, by Executive director.	Reports to board are informal made by staff.	ED and key staff make formal (written) reports to board.	Staff report to committees and committee chairs report to the board.	Executive board reports to full board.
Functioning & Role	M	Board not involved with administrative detail, or may be viewed as "rubber stamping" ED position.	Board may become highly involved with administrative detail or may be viewed as micro-managing agency.	Board involved with internal policies and procedures but may not have a clear understanding of the difference between policy and management.	Board focus is divided between internal and external policy issues. Mostly a tactical focus with some strategic focus emerging.	Board work is primarily external and strategic in nature. Tactical issues typically handled by staff.
	N	No committees - work completed by founder and volunteers.	No committees - members are a "hands on board' completing many assignments as volunteers.	Board committees for some functions like monitoring and reviewing financial status.	Standing & ad hoc committees that monitor agency functions and progress and report to full board.	Committees report to Executive Board which in turn reports to full board.

Features		Agency Stage of Development Revision 12.1 April 27, 2004				
		Stage 1 Initial, start-up or small agency	Stage 2 Advanced young or small agency	Stage 3 Established agency	Stage 4 Proficient agency	Stage 5 Highly productive agency
Executive and Board Relationship	*O*	ED leads board.	ED remains in board leadership role, but board may assert its oversight and authority; Board may conduct annual performance review of ED.	Board begins to differentiate between policy and management by delegating responsibility for the latter to ED.	Board differentiates between administration and policy roles; ED responsible for hiring staff; Board may approve nominees for high staff positions.	Evidence of board and staff collaboration; Staff may be empowered to make management decisions.
Policy	*P*	Board employs a contingency approach to solving management or policy problems as they arise.	Policy focus of the board is management (e.g., job descriptions, hiring and purchasing procedures, and financial controls etc.).	Board begins to differentiate between policy and management.	Board's focus is on broad policy issues; the ED is responsible for day-to-day operations.	Board interested in productivity, sustainability, and strategic issues of the organization; Internal and operational policies may be delegated to senior staff (e.g., HR policy).
Terms of service	*Q*	No specific term of office for board members.	Board term may be indefinite or function that way in practice.	Board serves for specified term and some rotation is expected but long-term membership is common.	Board serves fixed terms with policy for replacement to ensure continuity.	Board members may serve indefinitely or for agreed upon terms.
	R	ED recruits replacements for vacancies on board.	No policy for replacing non-functioning members or policy not applied.	Board has policy for filling vacancies and replacing non-functioning members.	Agency has mechanism for developing new members prior to service on board (e.g. service on advisory committee).	Vacancies rarely occur and when they do replacements have been identified and prepared well in advance.
Organization	*S*	Officers of convenience or may not have officers.	Board has officers but terms may be indefinite.	Officers with definite terms. Selection by board membership.	May have a provision for secession to offices (vice-chair becomes chair, finance committee chair becomes treasurer etc.).	May have a working executive board selected by full board with officers selected by the working board.
Decision-making	*T*	No formal process, usually ratification of ED recommendations.	Democratic decision-making usually by consensus.	Formal - internal or tactical focus. Staff may recommend but board action required.	Formal both internal and external focus with strategic elements emerging.	Formal and broadly strategic.
Uncalibrated indicators of stage change direction						
Change in Board Size		Board decreasing in size as a result of long-term attrition.	<----------	Expected direction of change	---------->	Board size increasing to bring new expertise or resources to agency.

Features		Agency Stage of Development Revision 12.1 April 27, 2004				
		Stage 1 Initial, start-up or small agency	Stage 2 Advanced young or small agency	Stage 3 Established agency	Stage 4 Proficient agency	Stage 5 Highly productive agency
Change in Board Participation		Active participation by a small group of dedicated members.	<----------	Expected direction of change	---------->	Active participation by most members
Change in Meeting Frequency		Meetings no longer held on a regular basis.	<----------	Expected direction of change	---------->	Increase in interval between regularly scheduled meetings.
Change in Board Composition		Majority of board members are charter members or have long terms on board.	<----------	Expected direction of change	---------->	Majority of members were not "founders" or on board early in agency history.
General Functioning of Board		More informal	<----------	Expected direction of change	---------->	More formal
Change in Board Focus		More tactical and short term focus.	<----------	Expected direction of change	---------->	More strategic and long term focus
Change in Board Tenure		Founder/Builder	<----------	Expected direction of change	---------->	Conserver
2. Financial Resources Maturity	*A*	**Unstable, unpredictable or insecure funding**	**Limited funding, more stable than stage 1 agency**	**Established agency funding where organization has some control over its own destiny**	**Stable sources of income or funding**	**Highly stable funding**
	B	Opportunistic funding (like that resulting from a new or emerging policy concern).	Contingency funding for activities or services but not long term (e.g. summer youth activities).	Initiative funding - Project funding derived from services related to an on-going policy priority like violence prevention or early childhood development.	Niche funding - Agency funded through on-going source like United Way or local funding authority (e.g. mental health, housing or juvenile justice).	Institutionalized funding for at least a core of services or programs (e.g. Funding is a line item on a state of local authority budget.
	C	Start-up or seed money only or short term funding.	Projects or activities funded on a one-time basis (funding may span several years).	Funding may be for specified periods but contains provision for renewal or continuation.	Core funding on-going, may have provision for periodic review or renewal.	Core service funding is highly stable. May be supported by spin-off services or products.

Features		Agency Stage of Development Revision 12.1 April 27, 2004				
		Stage 1 Initial, start-up or small agency	Stage 2 Advanced young or small agency	Stage 3 Established agency	Stage 4 Proficient agency	Stage 5 Highly productive agency
	D	Agency funding may not support core services completely.	Evidence of cross-subsidization of functions - management functions may rely on service providers taking on more than one job. Social work functions are not covered by fees, etc.	Evidence of cross-subsidization of services. Services that fit well with mission may rely on more profitable services that fit mission less well.	Core services generate surpluses in excess of expenses. Or agency receives some funding from investment, reserves, endowments, or capital assets.	Agency services and products generate sufficient revenue in excess of operating expenses to permit regular investment in new service development or innovation.
	E	No provision for recovering agency overhead cost.	Funding does not include sufficient overhead to cover actual costs.	Funding includes provision for support of indirect costs of administering the project or program.	Overhead supports infrastructure large enough to extend management functions beyond direct project administration.	High overhead supports a highly productive array of organizational functions.
Closeness of fit to mission	*F*	Agency may modify mission to more closely fit with priorities of available funding opportunities.	Activities of funded projects may not fit agency mission closely. Considerable reliance on cross subsidization to finance mission related activities.	Less reliance on revenues from poorly fitting activities to support closely fitting activities. Cross subsidization used primarily to develop a more comprehensive array of related services.	Non- or less mission related activities are part of a development plan or effort to build infrastructure or expand services.	Most activities are closely related to mission.
Fund raising	*G*	May be crisis oriented (e.g. to pay the rent or meet payroll.	Less crisis oriented, crises tend to be related to cash flow mechanisms.	Few crises unless the environment shifts in a major and unpredictable way.	No crises, strategic and long term planning.	Long term, strategic, developmental, and predictive of environmental needs.
	H	Responsibility of the ED or founder.	Responsibility of ED and staff.	Responsibility of ED, CFO and staff reporting to finance committee of Board.	Responsibility of Board and staff. Proposals usually developed by R&D or planning staff.	Board makes significant contribution to fundraising.
	I	Based on personal appeal and negotiated or non-competitive proposals.	New and renewing proposals or contracts some of which may be competitive.	Mostly relies on ongoing or renewing contracts and grants.	Relies on a mix of fundraising activities and investments.	Fields a full range of development activities and relies on product innovation and development to sustain mission.
Financial Vulnerability	*J*	Highly vulnerable, environmental stress may result in agency suspending most, if not all services.	Vulnerable, stress will be accompanied by reduction in some services.	Stable, environmental stress will not cause immediate service reduction.	Very stable, resources available to withstand stress without service reduction.	Highly stable, anticipates environmental stress in advance and makes strategic adjustments.
A. Equity Balance	*K*	Agency has no equity.	Agency has little or limited equity.	Agency developing equity.	Agency has some equity and implemented plan to enhance equity.	Agency has access to substantial equity balances.

Features		Agency Stage of Development Revision 12.1 April 27, 2004				
		Stage 1 Initial, start-up or small agency	Stage 2 Advanced young or small agency	Stage 3 Established agency	Stage 4 Proficient agency	Stage 5 Highly productive agency
B. Revenue Concentration - Diversity of Sources	*L*	Revenue from single IRS 990 source.	Generates program service income and/or contract or grant income.	Generates income from three 990 categories with largest portion coming from program service revenue.	Generates income from four IRS 990 categories but not well balanced among them.	Balanced portfolio with revenue from four of the five IRS 990 categories.
C. Revenue Concentration - Diversity of Funders	*M*	Income from a single funding source.	Income from more than one funding source or more than one grant or contract from a single source.	Income from several funding sources - or if a single source income is a niche in which agency faces little of no competition (e.g. Head start, domestic violence)	Agency has multiple funding sources of a comprehensive array of services.	Agency has multiple and balanced funding sources from more than one array of services (e.g. may have an MIS division that serves as an outsource to other agencies, or a consulting division.
D. Operating margins	*N*	Negative operating margin and no surplus, may have start up funding or may not be able to survive environmental stress.	No operating margin and low surplus, environmental stress will result in reduction of services.	Operating margin may be highly variable from year to year but agency has resources to stabilize services it provides.	Stable operating margin permitting strategic investment in product or infrastructure development.	Surpluses stable and sufficient to permit investment in product forecasting, development, & testing. Also invests in staff and board development.
E. Administrative costs	*O*	Not established.	Low or not fully recovered.	Moderate and may be fully recovered but infrastructure limitations may reduce productivity.	High overhead supporting well developed infrastructure.	Highest overhead relative to other agencies in field. Supports a highly skilled and diversified administrative staff.
Short-Term Change Annual Operating Budget	*P*	Recent large increase in annual operating budget.	<----------	Reserve Capacity	---------->	Recent large decrease in annual operating budget.
Intermediate Term Change Annual Operating Budget	*Q*	Large increase in annual operating budget.	---------->	Reserve Capacity	<----------	Large decrease in annual operating budget.
3. Organizational Maturity						
3 A. Vertical and Horizontal Differentiation	*A*	**Unrelated activities**	**Services organized into projects**	**Agency organized into programs of service**	**Agency organized into services**	**Agency organized into service systems**
	B	A single project or activity; or two or more unrelated projects or activities.	Agency has one or more projects comprised of one or more activities.	Agency program comprised of two or more related projects or components (e.g. outpatient MH + a community housing project).	Agency has one or more services comprised of two or more related programs (e.g. services for the homeless may combine shelter with substance abuse or mental health treatment).	Agency operates a system of services comprised of two or more unrelated or complementary services.

Features		Agency Stage of Development Revision 12.1 April 27, 2004				
		Stage 1 Initial, start-up or small agency	Stage 2 Advanced young or small agency	Stage 3 Established agency	Stage 4 Proficient agency	Stage 5 Highly productive agency
	C	Agency may have unrelated activities (e.g. advocacy, neighborhood watch, community organization).	Unrelated projects may represent funding realities of the time.	Niche service emphasis and approaches emerge in the agency.	Breadth of core services expands and becomes more comprehensive; Funding source diversity and operating size afford better opportunities for staff retention.	Core services of the agency address comprehensive scope of its mission.
	D	May have a "breakthrough" activities (i.e. an innovation on service or approach that has attracted funding).	Agency mission and services may be modified to better fit funding opportunities and realities.	Agency developing recognized programs of services and regular funding.	Agency's core services become a service system with more diversified funding and more comprehensive service offerings.	Agency may have a service system organized into divisions.
	E	Agency may have difficulty providing continuity of services.	Agency able to provide services but with lower skill level since it tends to employ former clients for entry-level jobs.	Service provision stabilizes unless affected by sudden environmental shifts.	Agency able to provide more stable and longer term services even with environmental shifts by using better trained employees.	Agency is best structured to provide continuous services and disruptive crises are non-existent.
	F	Projects may be slow starting owing to lack of required staff or resources like space.	Projects are less slow in starting; Agency tends to employ former clients/ entry-level staff; ED is likely to also be a project director.	Projects start on time because of internal promotion policies and procedures, or it has arranged for an external personal to develop proposal with hiring contingency.	Projects start on time by training employees in advance of startup to assume positions and responsibilities.	Projects start on time (or ahead of schedule) because agency has funded initial activities from internal surpluses.
	G	Retention in the agency is a problem because salaries and benefits are low; More part-time than full-time employees due to insecure sources of funding.	Funding sources are a little more secure with more full-time than part-time employees; Agency is experiencing high turnover due to non-competitive salary and benefits; entry level staff may be younger and/or inexperienced.	More stable employment situation in part due to competitive wages and benefit structures.	Stable employment situation, low turnover, and more experienced employees due to attractive salary and benefits.	Highly stable because agency invests in career development form entry level and up, and has a policy for promotion from within.
Internal Communications	H	Informal internal communications.	Internal communication largely informal with more formal elements being introduced.	Formal communications employed.	Both vertical and horizontal communication rely on more formal communications and arrange-ments than with smaller less well developed organizations.	Agency will possess both formal and informal multi-channels of communication.

Features		Agency Stage of Development Revision 12.1 April 27, 2004				
		Stage 1 Initial, start-up or small agency	Stage 2 Advanced young or small agency	Stage 3 Established agency	Stage 4 Proficient agency	Stage 5 Highly productive agency
3. B Administrative / Service Role Differentiation	*I*	Insufficient resources to permit meaningful role differentiation.	Administrative and service level differentiation emerges with key staff serving in both administrative and service functions.	Greater differentiation with some senior staff who have no direct service provision responsibilities.	Clear role differentiation and support role specialization.	Full differentiation and specialization with a career ladder for developing staff from entry level to senior positions.
	J	ED and perhaps one other staff member have highly developed service skills, but may not possess essential administrative skill levels.	Service skills of senior staff are at higher level than their administrative skills.	Service skills and administrative skill levels are roughly equal, but with administrative skills being slightly more extensive.	Service directors have staff who possess greater service skills than themselves	Senior leadership has extensive expert services and administrative skills.
3 C. Administrative Specialization	*K*	May have an "everybody does everything" culture.	Key staff have both administrative and service responsibilities.	Executive Director and/or CFO will not have service responsibilities.	Agency will have ED/CFO plus a director of operations who administers services.	Full management team for internal operations.
	L	Financial and legal functions handled by board members or volunteers as pro bono contribution to agency mission.	Will usually have administrative functions distributed between two or more individuals, who may also have direct service provision obligations.	Staff in support of both service and administrative functions to make senior personnel more productive.	Staff positions available to support product development or improvement.	Identifiable or dedicated planning and research staff.
	M	Administrative and financial management tasks may be responsibility of a single individual or board member.	May have a director and senior staff member responsible for financial management. Both jobs may be part time.	Executive Director and/or CFO pursue higher order administrative tasks and no client support or service responsibilities.	Executive director and CFO are free from day-to-day operational responsibilities.	Management team free from day-to-day operations responsibilities, and are associated with higher level functions and planning activities.
	N	Few, if any, support and clerical staff.	Agency employs some clerical support staff and may outsource some functions (e.g., bookkeeping, payroll).	Agency expands internal infrastructure to preclude outsourcing functions.	Agency may selectively outsource some functions for economies of scale.	Agency outsourcers core business functions but decision is based on economy and productivity reasons.
3 D. Multiorganization / Collaborative Arrangements	*V*	Group forms and talks about common or shared problems.	Group agrees on governance and takes action on common problems.	Emergence of some common organizational infrastructure and centralized resources.	Established organizational infrastructure and distributed infrastructure among partners.	Spin-offs or mergers.

Feature	Agency Stage f Development Revision 12.1 April 27, 2004				
	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5

4. Internal Operations Maturity

4 A. Management Information Systems

Infrastructure					
Space/facilities					
Technology					
Equipment					
Staff size & type					
Role differentiation					
Task specialization					
Expertise					
Domain					
Functional					
Notes					

4 B. Financial Systems

Infrastructure					
Space/facilities					
Technology					
Equipment					
Staff size & type					
Role differentiation					
Task specialization					
Expertise					
Domain					
Functional					
Notes					

4 C. Human Resources Systems

Infrastructure					
Space/facilities					
Technology					
Equipment					
Staff size & type					
Role differentiation					
Task specialization					
Expertise					
Domain					
Functional					
Notes					

Feature	Agency Stage f Development Revision 12.1 April 27, 2004				
	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
4 D. Communications					
Infrastructure					
Space/facilities					
Technology					
Equipment					
Staff size & type					
Role differentiation					
Task specialization					
Expertise					
Domain					
Functional					
Notes					
4 E. Development					
Infrastructure					
Space/facilities					
Technology					
Equipment					
Staff size & type					
Role differentiation					
Task specialization					
Expertise					
Domain					
Functional					
Notes					
5 A. Core Service:					
Infrastructure					
Space/facilities					
Technology					
Equipment					
Staff size & type					
Role differentiation					
Task specialization					
Expertise					
Domain					
Functional					
Notes					

APPENDIX D

SABI and LIFP Grantees Profiled

Trinity Heath Center
Mary Davey
319 Maple Street
Perth Amboy, NJ 08861

Community Action Service Center
Lydia Santoni-Lawrence, Executive Director
116 North Main Street
P.O. Box 88
Heightstown, NJ 08520

The Crisis Ministry
The Revrend Sallay Osmer
61 Nassau Street
Princeton, NJ 08542

Sikora Center, Inc
Susan Smith, Executive Director
613-615 Clinton Street
P.O. Box477
Camden, NJ

Woman Aware
Jacquelyn Marich, Ececutive Director
250 Livingston Ave
New Burnswick, NJ 08901

HomeFront, Inc.
Connie Mercer, Executive Director
1880 Princeton Avenue
Lawrenceville, NJ 08648

Mobile Meals of Trenton
Barbara Smith, Executive Director
546 Bellevue Avenue
Trenton, NJ 08618

Girl Scouts of Delaware-Raritan, Inc.
Dianne Fairbanks
Executive Director/CEO
108 Church Lane
East Brunswick, NJ 08816

Committee for Hispanic Children and Families, Inc.
Elba Montalvo, Executive Director
140 West 22nd Street, Suite 301
New York, NY 10011

The Center for Grieving Children, Teens & Families
Robert Sheesley, Director
The Nelson Pavilion
Erie Avenue at Front Street
Philadelphia, PA 19134

Urban Tree Connection
Skip Wiener, Executive Director
5125 Woodbine Avenue,
Philadelphia, PA 19131-2404

Lifeties, Inc.
Mary Inzana
CEO and Founder
2205 Pennington Road
Trenton, NJ 08638

Genesis Counseling Center, Inc.
Gabriel L. Guerrieri, Executive Director
566 Haddon Avenue
Collingswood, NJ 08108

Elijah's Promise, Inc.
Lisanne Finston, Executive Director
18 Neilson Street

New Brunswick. NJ 08901

Eric B. Chandler Health System
Rosemary McAndrew, Executive Director
277 George Street
New Brunswick, NJ 08901

Oak Orchard Community Health Center, Inc.
David W. Fisher, CEO
300 West Avenue
Brockport, NY 14420

New Brunswick Tomorrow
Jeffrey Vega, President
390 George Street
New Brunswick, NJ 08901

Children's Health Environmental Coalition
Elizabeth H. Sword, Executive Director
P.O. Box 1540
145 Witherspoon Street
Princeton, NJ 08542

Renaissance Community Development Corporation
Sharon Tucker Brown , Executive Director
630 Franklin Blvd., Suite 102
Somerset, NJ 08873

Local Initiative Funding Partners Grantees

Cancer Services of Allen County
Dianne May, Executive Director
2925 East State Blvd.
Ft. Wayne Indiana 46805

Eastern Panhandle Free Clinic
Kaye Napolitano, Executive Director
1008 Martin Luther King Jr, Blvd
Charles Town WV 25414

Putnam County Rural Health Clinics
Donna McSpadden, Executive Director
1225 West White Oak Drive
Cookville, TN
38501

BIBLIOGRAPHY

Bertalanffy, L. von (1968). *General System Theory.* New York: George Braziller.

Blalock, Hubert M. Jr. (1972). *Social Statistics*, Second edition. New York: McGraw-Hill Book Company.

Blau, Peter M. (1963). *The Dynamics of Bureaucracy*, Second Edition. Chicago: University of Chicago Press.

Blau, Peter M. (1970). A formal theory of differentiation in organizations. *American Sociological Review* Volume 35, Number 2, 201-219

Boruch, R. F. & Wortman, P. M. (1979). Implications of educational evaluation for evaluation policy. In Berliner, D. C., editor, *Review of Research in Education 7*. American Educational Research Association, 309-361.

California Wellness Foundation (2001). Reflections on capacity building. Vol. 2, No.2, Printed from www.tcwf.org/reflections/ on 1/14/2002.

Campbell, Donald T., Stanley, Julian C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Editor) *Handbook of Research on Teaching*. American Educational Research Association Chicago: Rand McNally, 171-246.

Concato, John, Shah, Nirav, and Horwith, Ralph (2000). Randomized, controlled trials, observational studies, and the Hierarchy of research designs. *The New England Journal of Medicine.* Volume 342, Number 25.

Cook, Thomas D., and Donald T. Campbell (1979). *Quasi-experimentation: Design & Analysis Issues for Field Settings*. Boston: Houghton Mifflin Company, 1979.

Cook, Thomas D., Leviton, Laura C., Shadish, William R. (1985). Program Evaluation. In The Handbook of Social Psychology. 3d edition., Lindzey G, Aronson E (eds). New York, Random House.

Cooley, William W. (1978). Explanatory observational Studies: AERA State-of-the-art address, annual meeting of the American Education Research Association. *Educational Researcher*, October, 1978.

Crane, John A. (1978). The power of social intervention experiments to discriminate differences between experimental and control groups, in Thomas D. Cook, Marlyn L. Del rosario, Karen M. Hennigan, Melvin M. Mark, William M.K. Trochim: *Evaluation Studies Review Annual*, Volume 3. Beverly Hills: Sage Publications.

Cronbach, Lee J. and Meehl, Paul E. (1955). Construct validity in psychological tests. *Psychological Bulletin* 52, 4, 281-302

Crosby, Philip B. (1980). *Quality is Free: The Art of Making Quality Certain.* New York: New American Library.

DiMaggio, Paul J. & Powell, Walter W. (1983). The iron cage revisited: instutitional isomorphism and collective rationality in organizational fields. *American sociological Review* 48:147-160.

Donaldson, Stewart I. (2001). Overcoming our negative reputation: evaluation becomes known as a helping profession. *American Journal of Evaluaiton*. Vol. 22, No. 3, 355-361.

Downs, Anthony (1967). *Inside Bureaucracy: A RAND Corporation Research Study*. Boston: Little, Brown and Company.

Eberts, Randall W. (1997). The use of profiling to target services in state welfare-to-work programs: An expample of process and implementation. Kalamazoo: W. E. Upjohn Institute for Employment Research.

Evans, Charles H. and Ilstad, Suzanne T., Editors (2001). *Small Clinical Trials: Issues and Challenges*. Committee on Strategies for Small-Number-Participant Clinical Research Trials, Board on Health Sciences Policy, Institute of Medicine. Washington, D.C.: National Academy Press.

Fredericksen, Patricia, and London, Rosanne (2000). Disconnect in the hollow state: the prvotal role of organizational capacity in community-based development orgnaizations. *Public Administration Review, Vol. 60,* No. 3. 230-238.

Freeman, Howard E. (1971). The implementation of research investigations. In Richard

O'Toole (Editor), The Organization, Management, and Tactics of Social Research. Cambridge, Mass. Schenkman, Publishing Company pp213-249.

Hall, A. D. and Fagen, R. E. (1968). Definition of System in Walter Buckley (Editor) *Modern Systems Research for the Behavioral Scientist: A Sourcebook.* Chicago: Aldine Publishing Company.

Harrison, Michael I. (1987). *Diagnosing Organizations: Methods, Models, and Processes.* Applied Social Research Methods Series, Volume 8. Newbury Park: Sage Publications.

Hermann, John A., Sambrano, Soledad, Springer, J. Fred, Sale, Elizabeth (1998). Why dosage data is critical: an examination of methodological issues. Paper presented at the annual meeting of the American Evaluation Association. Chicago: November 6.

Heyns, Roger W. and Zander, Alvin F. (1953). Observation of group behavior. In Leon Festinger and Daniel Katz (Editors) *Research Methods in the Behavioral Sciences.* New York: Holt, Rinehart, and Winston.

Horst, Pamela, Nay, Joe N., Scanlon, John W., and Wholey, Joseph S. (1974). Program management and the federal evaluator. *Public Administration Review* Vol. 34, No. 4, 300-308.

Howard, Kenneth I., Kopla, S. Mark, Krause, Merton S., Orlinsky, David E. (1986). The dose-effect relationship in psychotherapy. *American Psychologist* , Vol. 41, No. 2, 159-164.

Iscoe, Ira and Harris, Lorwen C. (1984). Social and Community Interventions. *Annual Review of Psychology*, Vol. 35.

Katz, Daniel and Kahn, Robert L. (1966). *The Social Psychology of Organizations.* New York: John Wiley & Sons, Inc.

Kazanjian, Robert K. (1988). Relation of dominant problems to stages of growth in technology-based new ventures. *Academy of Management Journal.* Vol.31, No. 2, 257-279.

Kazanjian, Robert K. and Drazin, Robert (1990). A stage-contingent model of design and growth for technology-based new ventures. *Journal of Business Venturing, 5,* 137-150.

Lawrence, Paul R. and Lorsch, Jay W. (1969). *Developing Organizations: Diagnosis and Action*. Reading Mass.: Addison-Wesley Publishing.

Leinhardt, Gaea and Leinhardt, Samuel (1980). Exploratory data analysis: new tools for the analysis of empirical data. In David C. Berliner (Editor), *Review of Research in Education 8*. American Educational Research Association.

Lesgold, Alan (2003). Detecting technology"s effects in complex school environments. In Geneva Haertel and Barbara Means (Editors) *Rigorous Research Designs to Evaluate Technology Effects.* New York:Teachers College Press.

Lesgold, Alan (2000). Understanding the effects of technology-based enhancement of professional development. Paper commissioned by SRI International for the United States Department of Education.

Letts, Christine W., Ryan, William P., and Grossman, Allen (1999). *High Performance Nonprofit Organizations: Managing Upstream for Greater Impact.* New York: John Wiley & Sons, Inc.

Leviton, Laura C. (1994). Program theory and evalution theory in community-based programs. *Evaluation Practice*, Vol. 15, No. 1, pp.89-92.

Leviton, Laura C. and Schuh, Russell G. (1999): The importance of a discovery capacity in community-based health and human service program evaluation. In Joseph Telfair, Laura C. Leviton, and Jeanne S. Merchant. editors: *Evaluating Helath and Human Service Programs in Community Settings. New Directions for Evaluation* Number 83, 17-35. San Francisco: American Evaluation Association & Jossey-Bass Publishers.

Leviton, Laura C. (2001). Building Evaluation's Collective Capacity. American Evaluation Association Presidential address, *The American Journal of Evaluation*, Winter, 2001, 22 (1), 1-12.

Light, Paul Charles (1998). *Sustaining Innovation: Creating Nonprofit and Government Organizations that Innovate Naturally.* San Francisco: Jossey-Bass Publishers.

Lipsey, Mark W., Cordray, David S., and Berger, Dale E. (1981). Evaluation of a juvenile

diversion program: using multiple lines of evidence. *Evaluation Review* 5(3):283-306.

Lipsey, Mark W., Crosse, Scott, Dunkle, Jan, Pollard, John, and Stobart, Gordon (1985). Evaluation: the state of the art and the sorry state of the science. *New Directions for Program Evaluation* No. 27, 7-48.

Lipsey, Mark W. (1988). Practice and malpractice in evaluation research. *Evaluation Practice* Vol. 9 No. 4, 5-25.

Lipsey, Mark W. (1990). *Design Sensitivity: Statistical Power for Experimental Research.* Newbury Park: Sage Publications.

Lipsey, Mark W. (1997). What can you built with thousands of bricks? Musings on the cumulation of knowledge in program evaluation. In Debra J. Rog and Deborah Fournier (editors) *Progress and Future Directions in Evaluation: Perspectives on Theory, Practice and Methods*. *New Directions for Evaluation* No. 76, 7-23. San Francisco: Jossey-Bass Publishers.

Miles, Matthew B. and Huberman, A. Michael (1984). *Qualitative Data Analysis.* Beverly Hills: Sage Publications.

Mintzberg, Henry (1994). *The Rise and Fall of Strategic Planning: Reconceiving Roles for Planning, Plans, Planers.* Toronto: the Free Press.

Mosier, Charles I. (1947). A critical examination of the concepts of face validity. *Educational and Psychological Measurement* Vol. VII. Washington, D.C.

Passell, Peter (1993). Like a new drug, social programs are put to the test. *The New York Times Science* Tuesday, March 9, 1993.

Perrin, Burt (1998). Effective use and misuse of performance measurement. *American Journal of Evaluation,* Vol. 19, No. 3, pp. 367-379.

Prochaska, James O., DiClemente, Carlo C., and Norcross, John C. (1992). In search of how people change: applications to addictive behaviors. *American Psychologist* 47:1102-1114.

Prochaska, Janice M., Prochaska, James O., and Levesque, Deborah A. (2001). A transtheoretical approach to changing organizations. *Administration and Policy in*

Mental Health. Vol. 28, No.4, pp 247-261, March 2001.

Quinn, Robert E. and Cameron, Kim (1983). Organizational life cuycles and shifting criteria of effectiveness: some preliminary evidence. *Management Science* Vol 29, No. 1, 33-51.

Reichardt, Charles S. (1994). Summative evaluation, formative evaluation, and tactical research. *Evaluation Practice* 15(3):275-281.

Reifler, Burton V. (2000). The role of technical assistance in philanthropy. Report to the Robert Wood Johnson Foundation. September 2000 (Revised in August, 2001).

Riecken, Henry W. (1976). Social experimentation. In Clark C. Abt (editor) *The Evaluation of Social Programs.* Beverly Hills: Sage Publications.

Robert Wood Johnson Foundation (2001). Budget Preparation Guidelines.

Rutman, Leonard (1980). *Planning Useful Evaluations: Evaluability Assessment.* Beverly Hills: Sage Publications.

Scheirer, Mary Ann (1987). Program Theory and Implementation Theory: Implications for Evaluators. In L. Bickman (editor) *Using Program Theory in Evaluation.* New Directions for Program Evaluation, No 33. San Franscio: Jossey-Bass, 59-76.

Scheirer, Mary Ann and Thayer, Collette (2000). Building organizatoinal capacity: likking literature and practice. Paper presented at the annual meeting of the American Evaluaton Association, Honolulu, November, 2000.

Schuh, Russell G. (1998). Developing successful demonstration programs. Paper presented in the workshop Byrdsong, Rashad, Schuh, Russell G. and Marin, Robert A Community-University Collaboration to Prevent Youth Violence a the annual conference of Community-Campus Partnerships for Health (CCPH), Pittsburgh: April 26.

Schuh, Russell G. and Leviton, Laura C. (1999). Creaming: peeking into the black box from a dose/response perspective. Presented at the annual meeting of the American Evaluation Association, Orlando, Fl. November 4, 1999.

Schuh, Russell G., Thompson, Kenneth S., and Lazar, Meryl A. (1999). Welfare to Work: Outcome Expectations vs. Mental Health Status. Interactive poster presentation at the American Psychological Association and National Institute for Occupational Safety and

Helath conference WorkStress & Health99: Organization of Work in a Global Economy. Balitmore, MD, March 12.

Schuh, Russell G., Leviton, Laura C. and Stagg, Vaughn (2000). Some Hidden Costs of Collaboration with Community Based Organizations. Paper presented at the annual meeting of the American Evaluation Association, Honolulu, November 3, 2000.

Scriven, Michael (1967). The methodology of evaluation. In Tyler, Ralph W., Gagne, Robert M., Scriven, Michael (Editors) *Perspectives of Curriculum Evaluation.* Chicago: Rand McNally & Company.

Sechrest, L., West, S. G., Phillips, M. A., Redner, R. and Yeaton, W. H. (1979). Some neglected problems in evaluation research: strength and integrity of treatments. In L. Sechrest, S. G. West, M. A. Phillips, R. Redner, and W. H. Yeaton (Eds.) *Evaluation Studies Review Annual* Vol. 4. Beverly Hills: Sage Publications.

Sechrest Lee and Yeaton, William H. (1981). Assessing the effectiveness of social programs: methodological and conceptual issues, in S. Ball (editor) *New Directions for Program Evaluation: Assessing and Interpreting Outcomes,* No. 9. San Francisco: Jossey-Bass.

Senge, Peter M., (1990). *The Fifth Discipline: The Art & Practicc of the Learning Organization.* New York: Currency Doubleday.

Shadish, William R. Jr., Cook, Thomas D., Leviton, Laura C. (1991). *Foundations of Program Evaluation.* Newbury Park, Sage Publications.

Shadish, William R. Jr., Cook, Thomas D., and Campbell, Donald T. (2002). *Experimental and Quasi-experimental Designs for Generalized Causal Inference.* Boston: Houghton Mifflin Company.

Shafritz, Jay M. and Ott, J. Steven (1987). *Classics of Organization Theory,* Second Edition. Pacific Grove, Ca.: Brooks/Cole Publishing Company.

Simon, Judith Sharken with J. Terence Donovan (2001). *The Five Life Stages of Nonprofit Organizations: Where You Are, Where You're Going, and What to Expect When You Get There.* St. Paul: The Wilder Publishing Center.

Stevenson, John F., Florin, Paul, D, Mills, S., Andrade, and M. (2000). Influencing and measuring evaluation capacity in human service organizations. Paper presented at the

annual meeting of the American Evaluation Association, Honolulu, November 4, 2000.

Suchman, Edward A. (1970). Action for what? A critique of evaluative research, in O'Toole, Richard (editor) *The Organization, Management and Tactics of Social Research* Cambridge, Mass: Schenkman Publishing Co.

Test, Mary Ann (1981). Effective community treatment of the chronically mentally ill: What is necessary? *Journal of Social Issues* Vol. 37 No. 3, 1981 71-84.

Tukey, John W., Gilbert, John P., and Mosteller, Frederick (1976). Steady social progress requires quantitative evaluation to be searching. In Clark C. Abt (editor) *The Evaluation of Social Programs*. Beverly Hills: Sage Publications.

Tukey, John W. (1977). *Exploratory Data Analysis*. Reading, Ma.: Addison-Wesley Publishing Company.

Vinzant, Douglas H. and Janet C. Vinzant (1996). Startegy and organizational capacity: finding a fit. *Public Productivity & Management Review,* Vol. 20 No 2, December 1996, 139-157.

Voss, James F., Terry R. Greene, Timothy A. Post, and Barbara C. Penner (1983). Problem-solving skill in the social sciences. In Gordon H. Bower (editor) *The Psychology of Learning and Motovation: Advances in Research and Theory* Volume 17. New York: Academic Press.

Wallace, Charles J. and Liberman, Robert Paul (1985). Social skills training for patient with schizophrenia: a controlled clinical trial. *Psychiatry Research,* 15:239-247.

Weiss, Carol H. (1972). *Evaluation Research: Methods of Assessing Program Effectiveness*. Englewood Clifts NJ: Prentice-Hall, Inc.

Wholey, Joseph S., Nay, Joe N., Scanlon, John W., and Schmidt Richard E., (1976). If you don't care where you get to, then it doesn't matter which way you go. In Clark C. Abt (editor) *The Evaluation of Social Programs*. Beverly Hills: Sage Publications.

Wholey, Joseph S. (1987). Evaluability assessment: developing program theory. In Leonard Bickman (editor) *Using Program Theory in Evluation, New Directions for Program evaluation,* no.33. San Francisco: Jossey-Bass, Spring:77-92.

Wholey, Joseph S. (1994). Assessing the feasibility and likely usefulness of evalution in Joseph S. Wholey, Harry P. Hatry, Kathryne E. Newcomer, (editors) *Handbook of Practical Program Evaluation.* San Francisco: Jossey-Bass Publishers.

Amherst H. Wilder Foundation (2000). Strengthening Urban communities: capacity building. Printed from foundation website www.wilder.org/suc/capbuild.html on June 2, 2000.

Wiley, David E. (1991). Test validity and invalidity reconsidered. In Richard E. Snow and David E. Wiley (editors) *Improving Inquiry in Social Science: A Volume in Honor of Lee J. Cronbach*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Yeaton, William H. and Sechrest, Lee (1981). Critical dimensions in the choice and maintenance of successful treatments: strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology.* Vol. 49, No. 2, 156-167.

Yeaton, William H. (1985). Using measures of treatment strength and integrity in planning research. *New Directions for Program Evaluation* No. 27 49-61.

Zaltman, Gerald (1979). Knowledge utilization as planned social change. *Knowledge: Creation, diffusion, Utilization.* Vol. 1 No. 1, September 82-105.