

NATIONAL
CENTER FOR
SERVICE
INTEGRATION

RESOURCE
BRIEF

10

CHILD AND FAMILY POLICY CENTER

Funding What Works:

Exploring the Role of Research on Effective Programs and Practices in Government Decision-Making

Charles Bruner
Mark Greenberg
Cynthia Guy
Michael Little
Lisbeth Schorr
Heather Weiss

with

Ron Cowell
Tom Gamble
Christina Groark
Clay Yeager



This publication is a joint product of the National Center for Service Integration Clearinghouse and the Center for Schools and Communities.



National Center for Service Integration

The National Center for Service Integration (NCSI) was established in 1991 with a grant from the U.S. Department of Health and Human Services. Its mission is to stimulate and actively support comprehensive, community-based service integration activities by serving as an information clearinghouse of documents programs, and materials related to service integration.

In June, 1995, the Child and Family Policy Center assumed responsibility to produce and disseminate NCSI publications. Although federal funding for NCSI has ended, NCSI continues to produce resource briefs, guidebooks, working papers, and other publications on issues communities and states face in developing more comprehensive, community-based service systems. The work of NCSI has been supported by various foundations, including the Annie E. Casey Foundation. NCSI partnered with the Center for Schools and Communities to host the Funding What Works Symposium and co-produce this publication. Charles Bruner, Director of the Child and Family Policy Center and the NCSI Clearinghouse, edited *Funding What Works*.



Center for Schools and Communities

Established in 1988, the Center for Schools and Communities is committed to improving outcomes for children and families through training, technical assistance, program evaluation, research and resource development. The Center focuses on initiatives related to education, violence prevention, and community and family, funded primarily by the Pennsylvania Departments of Education, Health, and Public Welfare.

Acknowledgements to Lynn Cromley, Director and Tom Robel, Special Projects Manager, who took the lead in organizing the Funding What Works Symposium as part of a larger Pennsylvania Conference on Building Partnerships, sponsored by the Pennsylvania Department of Public Welfare.

National Center for Service Integration
c/o Child and Family Policy Center
218 - 6th Avenue, Suite 1021
Des Moines, IA 50309
515.280.9027
Fax: 515.244.8997

Center for Schools and Communities
1300 Market Street
Lemoyne, PA 17043
717.763.1661
Fax: 717.763.2083

CONTENTS

Introduction: Funding What Works

- 1 CHAPTER 1
The Role of Evidence in Improving Outcomes for Children
Lisbeth Schorr
- 7 CHAPTER 2
Research-Based Programs and Issues of Implementation
Mark Greenberg
- 19 CHAPTER 3
Research and Policy: A Cross-National Perspective
Michael Little
- 23 CHAPTER 4
Making Progress: Learning and Public Accountability
Heather Weiss
- 31 CHAPTER 5
Developing Initiatives: A Foundation Perspective
Cynthia Guy
- 36 CHAPTER 6
At the Ground Level: State and Community Perspectives
*Tom Gamble, Ron Cowell, Chris Groark, and
Clay Yeager*
- 40 CHAPTER 7
**Summing Up: Points of Consensus on Using Research
In Government Decision-Making**
Charles Bruner
- 43 **Glossary of Terms**
- 49 **About the Authors**

Introduction: Funding What Works

There is increasing emphasis by state governments on developing accountability systems based upon achieving results. In education and human services, these results generally are defined in terms of improvements in measurable indicators of child or family well-being. This has led to identifying a set of child and family outcomes and indicators that can be measured throughout the state and over time. It also has led to efforts to construct new evaluation, financing, and quality assurance systems that are based upon achieving these outcomes.

In addition, this has led to efforts to use what we know in making decisions on what programs and services to fund – to draw from social service research on effective practices. This starts with identifying programs, principles, and processes that research has demonstrated produce these positive outcomes and then using these to inform policy-making and, in particular, new program funding decisions. *One of the emerging debates in the field is the extent to which state policies and new program funding decisions should focus on replicating proven research-based programs based upon a clinical research model or should take a broader focus in making program-funding decisions.*

In the research world, there is no consensus upon a single research methodology for determining which programs or practice principles meet a standard of proven success. While randomized controlled trials generally are considered the “gold standard” of research proof of a program’s efficacy and multiple site

studies with equivalent results across sites generally are considered as proof of replicability and broad-based effectiveness, only a few human services programs have been subject to such rigorous research.

Further, randomized controlled trials are most applicable to programs with quite specific and discrete goals. They are less applicable to more holistic, client-directed, and community-based programs (which may produce different outcomes for different subjects and even produce community-wide results that extend beyond their specific subjects). They can even violate the philosophy of certain programs that are predicated on being voluntary and inclusive in their acceptance of participants and can compromise those programs’ impacts as community builders.

Some researchers also have identified a broader set of attributes of effective practice — holistic, strength-based, client-directed, community-based — that, in combination, have produced evidence of significant success, particularly for vulnerable children and families who have not responded well to traditional services. Alternative methodologies have been proposed and employed to assess the effectiveness of these attributes, but these have sometimes been critiqued as lacking the vigor and scientific validity of randomized trials.

The presentations in this *Resource Brief* tackle these and other issues involved in moving toward outcome accountability in “funding what works.” The presentations

were part of a special Research Symposium in Harrisburg, Pennsylvania, co-sponsored by the National Center for Service Integration and the Center for Schools and Communities.

The symposium was conducted in Pennsylvania on May 18th, 2001, and was timely because Pennsylvania has begun to tackle this issue. Pennsylvania recently established a special funding pool to invest in the replication of research-based programs. At the same time, the state long has supported systems reform efforts through Family Service Systems Reform (FSSR) and Family Center (FC) grant programs that have been based upon incorporating attributes of effective practice into existing and new, more preventive programs. It is seeking to incorporate outcome accountability and the use of research findings into all of these state investments.

Lisbeth Schorr led off the symposium by making the case for use of a broader range of evidence than that obtained from randomized clinical trials in making funding decisions. She further articulated a set of attributes of effective practice and their own body of evidence supporting their effectiveness.

Mark Greenberg argued for a “multiplicity” of research approaches, but contended that randomized controlled trials offered the greatest certainty of program impact and should be employed where feasible. He also spoke to the need to do more than fund research-based programs and to give specific attention to implementation efforts to assure the fidelity of their replication.

Michael Little offered a cross-national perspective and warned against examining programs against an all-or-nothing standards. Rather, research should look for more than statistical significance and address the questions, for whom, by how much, and under what conditions. Little further indicated that, rather than starting from a program base, a more promising approach is to start from a child needs base.

Heather Weiss noted the importance of developing an infrastructure to support evaluation and continuous learning as part of organizational and program development. She contended that outcome accountability is part of the political landscape, and, unless practitioners and researchers develop strategies to continuously improve practice through use of outcome data, funders may conclude that little works and limit their funding to find solutions.

Cynthia Guy provided one foundation’s perspective, both as a funder and a consumer of research. She argued for a strong recognition that funders need to support innovation in areas where there are not research-based program models but there exist important social concerns.

A panel of Pennsylvania respondents — **Tom Gamble, Ron Cowell, Chris Groark, and Clay Yeager** — then offered their perspectives on the implications for state policy and practice.

Charles Bruner, who moderated the panel, concluded by offering his summary of the consensus points from panelists and respondents.

This *Resource Brief* provides all these presentations, edited by the individual presenters. While different presenters used somewhat different terminology, an effort was made to use a common language in discussing research and evaluation and distinguishing between different research methodologies and findings. The **Glossary of Terms** at the end of this *Resource Brief* provides a set of definitions of terms for use by participants at the symposium.



NATIONAL CENTER FOR SERVICE INTEGRATION PUBLICATIONS

In addition to this document, the National Center for Service Integration has produced nine other resource briefs, five working papers, and three guidebooks on topics of interest to states and communities involved in comprehensive service reform.

RESOURCE BRIEFS (**\$4 each unless otherwise noted**)

- 1 *So You Think You Need Some Help? Making Effective Use of Technical Assistance*, by Charles Bruner
- 2 *Charting a Course: Assessing a Community's Strengths and Needs*, by Charles Bruner, Karen Bell, Claire Brindis, Hedy Chang, and William Scarbrough
- 3 *Who Should Know What? Confidentiality and Information Sharing in Service Integration*, by Mark I. Soler and Clark M. Peters
- 4 *Getting to the Bottom Line: State and Community Strategies for Financing Comprehensive Community Service Systems*, by Charles Bruner and Frank Farrow
- 5 *Getting Started: Planning a Comprehensive Service Initiative*, by Carolyn Marzke and Deborah Both
- 6 *Making it Simpler: Streamlining and Integrating Intake and Eligibility*, by Allen Kraus and Jolie Bain Pillsbury
- 7 *Making a Difference: Moving to Outcome-Based Accountability for Comprehensive Service Reforms*, by Nancy Young, Sig Gardner, Soraya Coley, Lisbeth Schorr, and Charles Bruner
- 8 *Wise Counsel: Redefining the Role of Consumers, Professionals, and Community Workers in the Helping Process*, by Charles Bruner, Edgar S. Cahn, Audrey Gartner, Robert P. Giloth, Toby Herr, Jill Kinney, Janice M. Nittoli, Frank Riessman, Margaret Trent, Yolanda Trevino, and Suzanne L. Wagner (\$8 each)
- 9 *Resident Experts: Supporting Neighborhood Organizations and Individuals in Collecting and Using Information*. Veronika Kot and Charles Bruner

WORKING PAPERS (**\$4 each**)

Beyond the Buzzwords: Key Principles in Effective Frontline Practice, by Jill Kinney, Kathy Strand, Marge Hagerup, and Charles Bruner

Steps Along an Uncertain Path: State Initiatives Promoting Comprehensive, Community-Based Reform, by Charles Bruner, Deborah Both, and Carolyn Marzke

Realizing a Vision for Children, Families and Neighborhoods: An Alternative to Other Modest Proposals, by Charles Bruner, with foreword by Douglas Nelson and commentary by Otis Johnson

Reinventing Common Sense, by Judith Levey, with introduction by Charles Bruner

Social Services Systems Reform in Poor Neighborhoods: What We Know and What We Need to Find Out, by Charles Bruner

GUIDEBOOKS (**\$12 each**)

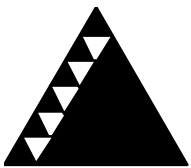
Defining the Prize: From Agreed-Upon Outcomes to Results-Based Accountability, by Charles Bruner

Valuing Diversity: Practicing Inclusion, by Hedy Nai-Lin Chang and Charles Bruner

Getting to the Grassroots: Neighborhood Organizing and Mobilization, by Charles Bruner and Maria Chavez

Publications may be ordered online at www.cfpciowa.org

or by writing: National Center for Service Integration, c/o Child and Family Policy Center, 218 - 6th Avenue, Suite 1021, Des Moines, Iowa 50309.



National Center for Service Integration
c/o Child and Family Policy Center
218 Sixth Avenue
1021 Fleming Building
Des Moines, IA 50309-4006

CHAPTER 1

The Role of Evidence in Improving Outcomes for Children

Lisbeth Schorr

I am delighted to be part of this timely symposium which promises to illuminate the debate on what constitutes credible research on programs and policies to improve outcomes for children and families. I want to contribute to this debate by making two essential points.

First, I think we have to distinguish between research-based and evidence-based decision-making in assessing whether programs work. Second, I will argue that in identifying what works, it is not enough to identify effective programs. We also have to identify the attributes that are essential to their success and the infrastructure that is required to support and combine and sustain what works.

We can all join in celebrating the impulse that is so prominent here in Pennsylvania — to move away from the purely anecdotal or intuitive approach to deciding what defines success and to move toward a much more rational, thoughtful, systematic, and rigorous way of determining what deserves public support. To the extent that Pennsylvania's emphasis on funding research-based programs means getting away from how little we had to go on in earlier days, I think it represents an enormous leap forward. It's a big leap forward from the situation we faced only a decade ago, illustrated by the survey done by a major foundation of its youth development grantees. They were asked what they were doing and why they were doing it. Of the twenty-nine grantees surveyed, twenty-eight responded

by saying either, "Well, it seemed like a good idea at the time," or "It's what we could get funding for." Only one of the twenty-nine even attempted to make the connection between the program's activities and what it was trying to achieve for youth.

Today, we are getting away from funding people whose applications basically say, "We are well-intentioned and we are trying hard, so you should give us the money." We are no longer content to simply fund good intentions, and we are becoming much clearer about what we are trying to achieve. We are getting much better at defining our hypotheses about the connections between our activities and the outcomes we are striving for. We have become much more insistent on asking for evidence that we will achieve those outcomes.

But what happens when we insist that those connections be confirmed by research? As long as research is considered credible only if it meets traditional conventions that come out of the biomedical sciences, I think we will be poorly served. I believe that, in efforts to seem "scientific," evaluators of social programs sometimes have been bullied into testing social interventions the way the FDA tests drugs, using biomedical research methods that are not appropriate to many social programs.

A Brookings Institution report recently insisted that human services reforms should be judged legitimate or illegitimate by the application of randomized trials in which, I

When we rely only on randomized experiments we miss too much information that is crucial to improving complex policies and programs.

quote, “One person gets the pill and the other the placebo.” I believe that this reveals a massive misunderstanding of the most promising kinds of social programs that exist today, which are often quite complex and are adapted to individual conditions and needs. Evaluating these complex social programs is not like testing a new drug.

The interventions that are now needed to rescue inner-city schools and strengthen families and rebuild neighborhoods are not stable chemicals that are manufactured according to standardized specifications and administered in standardized doses to passive recipients. Promising social programs often are complex efforts with multiple components that require constant mid-course correction, the active involvement of committed human beings, and flexible adaptation to local needs and strengths, to lessons learned, and to changing circumstances. It is the very nature of the most promising programs that makes them almost impossible to evaluate the way we evaluate drugs.

It is true that when a social program is sufficiently circumscribed and standardized, it can be evaluated by FDA methods, using randomized trials. When it is, then we can have “research-based” proof that it is effective. Let us think about what we lose, however, when we limit our funding only to interventions for which we have such proof.

At a recent meeting of the Brookings

Institution’s Children’s Roundtable an economist presented a paper on what we knew about the effectiveness of after-school and youth development programs. He concluded that we know so little about what works in that domain, that there should be no public funding going to those programs until we’ve done more research.

I was one of the respondents to that paper and I listed some of the things that I believe we know about after-school and youth development programs that work. I was able to draw on a large body of evidence, including evidence from practice, systematically analyzed — evidence that was outside the domain of my economist colleague because it did not include *proven causal connections*. When we rely only on randomized experiments we miss too much information that is crucial to improving complex policies and programs that profoundly affect the well-being of children and families most at risk today.

Consider the example of the Pittsburgh Early Head Start program, established in four centers in three Pittsburgh neighborhoods. It involves multi-purpose home visiting, combined with group activities for both children and families. It works with all types of child care providers (including child care centers, family day care homes, neighbors, and relatives). The program not only provides toys, educational and family support services, but it also provides help for parents who are applying for jobs or are trying to get social services. Each family has

We need to be able to use a much richer set of clues and evidence to determine what programs are worth supporting, and which are not, and which need to be revised.

regular contact with a team, including child development specialists, nurses, family advocates, and community organizers. Now you tell me how this multi-dimensional and individually tailored program can be evaluated using the experimental methods that the FDA uses. It cannot be done. And if we can't prove that it works, how do we know that it is promising? Because we have evidence that the pieces work, and by putting together what we know from theory and practice we can predict that the pieces will work a lot better when they are put together.

I think we have to conclude that the biomedical research methodologies that provide “gold standard” proof in other contexts cannot provide sufficient evaluative evidence about many of our most promising interventions, with their many interactive and evolving components.

On the chart on the following page I have tried to lay out this mismatch between the gold standard of intervention and the gold standard of evaluation. On the left hand side are a list of attributes of programs that have come to be recognized as important in succeeding with children and families that too often have not been well-served by current programs and systems. These attributes contrast sharply with the program characteristics that lend themselves to evaluation by traditional biomedical methods, and the mismatch is clear. If we use only the elegance of the randomized trial to judge what is effective, we have such

a narrow funnel through which promising programs have to travel that we get very slim pickings at the bottom.

We need to be able to use a much richer set of clues and evidence to determine what programs are worth supporting, and which are not, and which need to be revised.

We also need to be able to pursue hypotheses, based upon our findings to date. For instance, we know that both research and practice have shown that the social isolation of families with infants and young children is a major risk factor for damage in later life. The research that has assessed most of the programmatic efforts to reduce this social isolation has not found a lot of strong effects. Most family support centers and home visiting programs — programs designed to help families become more connected and better raise their children — show relatively weak effects on child outcomes.

One of the reasons for these weak findings may be that the efforts to intervene that we have evaluated have been too circumscribed. We have not used the disappointing results from these evaluations to generate new hypotheses. We have not systematically focused on the possibility that we got those weak effects because a high proportion of the mothers that were visited were depressed or were drug abusers and the home visiting program alone was not enough of an intervention. If you were to combine that home visiting program with

**MISMATCH BETWEEN THE “GOLD STANDARD” OF INTERVENTIONS
AND THE “GOLD STANDARD” OF EVALUATIONS**

ATTRIBUTES OF EFFECTIVE INTERVENTIONS

**ATTRIBUTES ASSOCIATED WITH TRADITIONAL
“EVALUABILITY”**

significant front-line flexibility within established quality parameters



intervention standardized; discretion minimized

evolving -- in response to experience and changing conditions



intervention constant over time

intervention/program design reflecting local strengths, needs, preferences



intervention centrally designed, uniform across sites

intake/recruitment into program under local control within broad parameters



intake/recruitment centrally designed to permit random assignment

multiple components respond to children in family, peer, & neighborhood context



single factor, single sector

interactive components take into account health, social, educational needs



components clearly separable

emphasize continuing respectful relationships, other hard-to-measure attributes



readily measured inputs

implementers “believe in” the intervention



value-free implementation

the capacity to mobilize a mental health intervention or a drug treatment intervention when those were needed, or with being able to mobilize more intensive services with the children when those were needed, then you might have a much more promising intervention, if one much harder to evaluate.

The premise on which so much of the recent

evaluation work has been based is that unless one can establish the program’s impact with certainty, one must conclude that there is no evidence that the program is working. We have to reject that premise. To do otherwise would interfere with program and policy development that could ultimately lead us where to where we need to go to expand our knowledge base and get better results for children and families.

You can ask a much richer set of questions when you're willing to accept multiple ways of knowing and various levels of certainty about program impact.

When we combine what we know from theory, experience, and research, we can begin to identify not only programs that work but also the essential attributes that seem to emerge from most efforts to improve outcomes among high-risk populations. In addition, we can identify the infrastructure necessary to sustain these programs.

When we focus on the attributes that make programs effective, we can solve another problem that has bedeviled efforts to scale up and spread model programs. On the one hand, we all know and we all say that we cannot simply replicate, or clone, a successful model. We have to adapt it to meet local circumstances. But when adapting, we need to know we are retaining essential elements and are not undermining what works through the adaptations. One of the ways of knowing this is by figuring out what, indeed, are those essential elements. If we know that in some depth, our programming will be much better informed and more effective, whether the initial determination of program effectiveness was made through randomized trials or by other means.

New approaches are now being developed to take a more dynamic approach to building knowledge of what works. In my own efforts to learn more systematically from experience and research, I have been exploring a particular approach that does not start with a review of published research. Instead, it starts by asking people who are deeply involved in the field about what

they, on the basis of their research and their practice experience, believe works. This approach produces rich discussions that are not just about programs, but about program attributes, staff qualities and problems of implementation in local contexts. After organizing the results of those discussions, we are looking at the research to determine which of the expert observations are indeed supported by traditional research and, for the rest, why there is not a research base. Has the particular issue simply not been researched? Have results from existing research shown no or few effects? If so, was the research based on programs that did not incorporate the essential elements of the program? Are new research designs needed that would be a better fit with the essential elements of the promising programs? You can ask a much richer set of questions when you're willing to accept multiple ways of knowing and various levels of certainty about program impact. When you get answers to these questions, you also get a much richer knowledge base from which to draw in shaping programs and policies.

By being more inclusive as to what counts as credible evidence, I believe we can create a knowledge base that moves the whole field away from oversimplified judgments about program success and failure. This will build a richer, more complex body of information about strategies that are plausible, promising, or proven. Practitioners, program designers, and communities then can make use of the lessons learned from both research and experience to construct

By being more inclusive as to what counts as credible evidence we can create a knowledge base that moves the whole field away from oversimplified judgments.

stronger theory and more effective interventions. By applying intelligence and judgement to better understand the accumulation of existing research and experience and by building on that in the future, we can construct something that will be vastly stronger in getting us to the outcomes we seek.

I hope that your commitment here in Pennsylvania to become more rational and systematic about determining what is worth funding will lead you to become part of, and possibly leaders of, efforts to build this richer, more inclusive knowledge base.

CHAPTER 2

Research-Based Programs and Issues of Implementation

Mark Greenberg

The goal of my talk today is to examine the strengths and limitations of different research methods to study program effectiveness. Like Lee Schorr, I will make the case for a “multiplism” of research models and I very much agree with her statement, “We should stop treating methodological issues as religious quarrels” (Schorr & Yankelovich, 2000). Further, I also agree with her statement that we should use a “rich set of clues” to understand and evaluate programs. However, I will also argue for using randomized clinical trials where possible, but also employing other techniques where more appropriate or feasible.

I might add that I find the distinctions between evidenced-based approaches vs. research-based approaches quite murky in this regard. Thus, I will refer to “empirically-validated programs” as those that have undergone a rigorous evaluation using either a randomized or quasi-experimental trial and where the programs have been found to significantly alter the developmental processes of children’s outcomes. Finally, I will then highlight the quality of program implementation as a critical determinant in replicating effective programs.

Before I begin to discuss different research methods, however, I want to make several broad points related to any form of program evaluation and assessment.

First, we do know quite a bit about the characteristics, or attributes, of effective programs, as have been summarized by Lee

Schorr, Joy Dryfoos, and others (Dryfoos, 1998; Schorr, 1997; Weissberg & Greenberg, 1998a). We know that programs that are more comprehensive, more flexible, and more responsive to the needs of participants are more likely to be effective.

Second, we know that it is important to view children in the context of broader ecologies — families, schools, neighborhoods, churches, and communities. We also know there are almost no quick fixes that actually work. To improve the quality of life of children and families we need to take a preventive and long-term approach that is developmental in nature and one that operates in coordination with these other forces in a child’s life. Effective programs need to be linked with other systems of support and intervention to ensure they can produce and sustain their impacts over time.

Third, effective programs generally are operated by people with a commitment and intensity to their work and a clear sense of mission. The first step in many prevention programs is building a solid relationship with the child or family, which requires establishing mutual trust. If program staff are not emotionally invested in the people they serve and do not establish those trusting relationships, the program design and structure itself will not produce results. Effective programs are based upon quality staff with effective models of training and ongoing technical assistance, and are not simply the result of a program curriculum

I will make the case for a “multiplism” of research models... However, I will also argue for using randomized clinical trials where possible.

or design.

With this as context, I want to examine the strengths and limitations of several different research methods that investigate whether programs are effective. It is also important to note that evaluation has not played a role as a driving force in the development and refinement of many existing programs.

The challenge we face, and why we are here today, is that, although there have been programs around for many years that have been supported and even touted as effective or promising, we do not know that much about their ultimate impact on children. I quote Doug Powell, a noted researcher in the family support field, who has stated:

The development and expansion of family support programs have far outpaced the availability of research information on program implementation and effectiveness. This is not a field where science has played a major role in informing practice. (Powell, 1994)

We are not talking here about randomized trial evidence versus other types of evidence. We are talking about any levels of evidence.

This is true for other program areas, as well. In spite of the enormous funding that they have recently received, after-school programs do not have a rigorous research base. There is little evidence that such

programs produce sustained impacts upon the students they serve, nor what types of after-school programs are effective with what types of students. We know little about how such issues of quality of staff, content of programs, or dosage (should children attend 5 days a week) of after-school programming impacts children’s academic or social outcomes. A further example at a broad-based level is that there is very little research evidence that Communities that Care (CTC), as a community planning process, produces impacts. It is only in the past few years that Pennsylvania has taken the lead in examining factors that influence the effectiveness of CTC (Greenberg, Feinberg, & Osgood; 2001).

These examples are used only as illustrations of program areas where we are proceeding with little research evidence to guide us. These programs all make good common (and theoretical) sense, and they are seen as promising approaches to address child and family needs. However, we have little evidence to draw upon to tell us they are producing their desired results.

This evidence can only be obtained through solid research. This has led the Institute of Medicine to argue that for the reduction of risk factors, the promotion of protective factors and the ultimate prevention of mental disorders:

There could be no wiser investment in our country than a commitment to foster the prevention of mental disorders and the

The RCT helps us to know unequivocally whether or not the program is effective.

promotion of mental health through rigorous research with the highest of methodological standards. Such a commitment would yield the potential for healthier lives for countless individuals and the general advancement of the nation's well-being. (Mrazek & Haggerty, 1994)

The Institute report goes on to state the following:

*The choice of research methodologies is a major issue in examining preventive interventions. It determines whether evidence is compelling. The **ideal** design is a randomized controlled trial.*

The Institute report does not state that the randomized controlled trial is the only research methodology that should be employed or that it is appropriate to all programs and interventions. It does, however, state that it can produce the most compelling evidence of program impact.

I want to expand upon this point as I discuss both the strengths and the limitations of randomized clinical trials (RCTs).

Simply put, the RCT helps us to know unequivocally whether or not the program is effective. That is only a part of what ultimately we need to know, but it does give us the most compelling evidence of program impact.

In using a pre-test and post-test model only, we cannot be certain that the impact is the

result of the program and not some other factor. Measured improvement from pre-test to post-test could be the result of normal growth and development that had no connection to the program, or it could be due to other forces operating during the same period, such as changes in the economy or in other services and supports in the community.

In using a comparison group drawn from a different population (such as a group from another community or a group not recruited into the program), we cannot be sure that different forces were not at work in that other community that produced the differences or that there was self-selection into the program by those most likely to achieve gains, without or without program participation.

Pre- and post-test studies that have either no control group or a “matched” control group design simply cannot provide the certainty of program impact that a randomized trial can provide; there are too many counterfactuals.

A single RCT showing evidence of program effectiveness only can provide certainty of program impact to the specific setting in which it was conducted. Additional trials in other settings that replicate the effects greatly increases certainty that the program will produce those effects in settings similar to the ones in which multiple trials were conducted. Second, replication of results greatly increases the confidence in the

As a program is replicated, with similar impacts, with diverse populations and in diverse settings, we gain more confidence that it is very broadly, if not universally, effective.

program's general effectiveness. There are a growing number of prevention programs that have met this test of effectiveness. As example, The Nurse Family Partnership Program (formerly called the Prenatal Nurse Home Visiting Program) has now shown replicated effects across separate trials (Olds et al., 1998).

Thus, there still are important caveats to relying upon randomized clinical trial results as the sole criteria for program funding, even when those results are available.

First is the issue of replication. We need to know under what conditions the program is effective and if this matches the conditions where the program will be implemented. The Colorado Blueprint series set multiple site replication as one of its standards for inclusion as an effective program. For instance, replication can establish whether the program is effective with different racial or ethnic groups, in both urban and rural settings, with persons of different levels of education and in different types of communities. For example, the Nurse Family Partnership Program has randomized trials showing its effectiveness in inner city as well as rural communities, and with several different ethnic groups. As a program is replicated, with similar impacts, with diverse populations and in diverse settings, we gain more confidence that it is very broadly, if not universally, effective.

Second is the issue of implementation. We

need to develop a set of training procedures to ensure that programs are being implemented with fidelity to the program model that was tested. Without fidelity of program implementation (see below for further elaboration of the components of implementation), there can be no expectation that program impacts equivalent to those established by randomized trials will occur. Even with Blueprint programs, we need to ensure the availability of adequate training and technical assistance to new program sites to ensure fidelity in program implementation.

The third issue is one of flexibility and adaptation. In very few instances can a program simply be replicated, without some adaptation to local conditions. Adaptations may include the integration of the program with other available services and supports. It may include adaptations in the "home" or location of the program within the community, both in terms of geography and agency, in order to assure that it is accessible and credible. It also may include variations in "social marketing" to the culture of different communities. At the same time this adaptation occurs, however, it is essential that there be fidelity to the underlying characteristics required to make the program effective.

The fourth issue is that of ongoing program evaluation and review. In the end, all programs require monitoring systems to gather data on changes they have produced in the individuals they have served. They

There is a real danger in using “best ideas” in the absence of effective research evidence...
...there have been a number of “best ideas” that have been shown to have minimal or no effects.

may not require local RCTs, but local replications do need to measure their impacts, which can be contrasted against the impacts achieved in programs that have been subject to more rigorous evaluations.

I hope I have established that, where possible, randomized trials provide the most compelling evidence for program effectiveness, but they are not sufficient by themselves to guarantee the effectiveness of new programs based upon that program model.

I also want to emphasize that RCTs are not the sole research tool that should be employed in testing programs, and that some programs, because of their structure, may not be capable of evaluation by randomized trials. Further, all programs, using RCTs or other models, should also be gathering qualitative data to better understand how and why their program succeeds or fails.

One of the greatest challenges in program evaluation is to figure out which type of study design will best fit the question. Further, we need to understand how to combine different methodologies together to gain more certain answers to questions of what works and to gather more support for further research and experimentation in the field.

Compared to RCTs, the “next best” design in producing compelling results are quasi-experimental designs. Although there is no randomization, if done with care, one can establish a comparison group, which has the

same characteristics as the control group. A good example of a quasi-experimental design is the research conducted by Art Reynolds of the Chicago Preschool Project (Reynolds et al., 2001). This project involved a study of over one thousand children who received Title I funding for parent and child services in pre-school through first grade that were compared with five hundred children from a very carefully selected comparison group in similar Chicago neighborhoods. The research has followed the children through high school and documented a 12% reduction in school dropouts among children in the project compared with the comparison group of children. The study also collected evidence to suggest that the most important elements for success were the very active involvement of parents and the strong overall focus on reading in the preschool period.

Quasi-experimental studies can be effective, and we can learn a great deal from them. They do require statistical controls to carefully match comparison groups with participants. If possible, randomized trials can later be employed to establish greater certainty of the results and begin the replication process.

There is a real danger, however, in using “best ideas” in the absence of effective research evidence. Unfortunately, there have been a number of “best ideas” that logically would seem to address a need. However, after careful research has been conducted, these programs have been shown

Research may be as important in learning what does *not* work as in finding out what does. If well used, we can avoid making old mistakes in our funding.

to have minimal or no effects. For example, because of substantial interest in and development of early childhood case management programs around the country, the federal government funded the Comprehensive Child Development Program (CCDP) in twenty-one communities, incorporated a randomized controlled research design for each program. Although averaging a program expenditure of \$14,000 per year per child for a two-year period, the research showed no overall effect on child development or parent functioning (St. Pierre, et al., 1997). We have learned a lot from that federal study. It has given us some insights into what might have been missing in the CCDP model that would need to be incorporated into subsequent efforts to achieve success. It also has warned us against replicating the CCDP model and expecting to achieve any real gains. Without this research, however, there would have been a great likelihood that CCDP-type programs would continue to be presented as effective programs and more broadly funded, despite their lack of evidence of impact.

In fact, the research may be as important in learning what does *not* work as in finding out what does. If well used, we can avoid making old mistakes in our funding and we can retool our “best ideas” into something more likely to succeed. We have a good deal of research about some “best ideas” that do not work. Some forms of home visiting have failed to show impacts, as have many job training programs. Despite continued support, we have convincing clinical trial

data that shows that Drug Abuse Resistance Education (DARE) does not reduce drug use (Ennett et al., 1994). Most programs aimed at smoking and tobacco prevention for children have not proved effective, nor have public relations campaigns and commercials (Lantz et al., 2000). Such research does not say that the problem is not an important one to address or that there is no way to address it, but it can help us redirect our funding and our energies to new program development and testing in order to develop effective approaches. For this, we do need good research that has a strong control or comparison group or other counterfactual.

Agencies, which develop and implement innovative programs almost uniformly believe that their programs work. I have never been to an agency that has developed an innovative program, which did not believe that their program worked or that it should not be widely expanded and disseminated. Although most agencies provide some data to justify their claims, most of this program evaluation data does not directly connect their claims to program impact. In fact, many innovative programs do not work, and we should expect that and learn from them.

In order to use research as a continuous learning tool, there needs to be a clear framework for accountability. This does not necessarily mean the use of RCTs, but it does require starting off by establishing program expectations and identifying and collecting the rich set of clues that we can

RCTs cannot be employed in all situations. Efforts to alter and coordinate large service delivery systems do not fit into a randomized controlled design.

collect that will help assess whether the expectations are being met.

Without good research evidence, programs are likely to take credit for any improvements that occur. Today, many programs take credit for a reduction in teen pregnancy, a reduction that has occurred throughout the country. Programs focusing upon abstinence and those focusing upon birth control awareness and responsible sexuality both take credit for this decline. They both could be right; they could both be neutral; it could be that neither of them are having a significant impact; but we really will not be able to determine what works for what teens until we have a better research base. That is the danger of using “best ideas” without careful evaluation design. We will not have a way of attributing, and sustaining, gains that may occur and we will not be able to know whether we should expand, continue, or stop those programs that are designed to address a real public need.

It should be recognized that RCTs cannot be employed in all situations. Efforts to alter and coordinate large service delivery systems do not fit into a randomized controlled design, although specific elements within them may. Randomly assigning individuals to a program that relies upon being inclusive and open to all community members may violate the program’s own tenets and one potential program strength. Specific example are the recent studies that have examine the effectiveness of systems of care

models for children’s mental health services in which comparison communities were used along with sophisticated statistical methods to control for community differences.

Some programs may have been contoured to a certain community due to history or politics or a specific precipitating event. What made that program effective may be highly individualized and not transferable, and a randomized trial will not be much help to furthering the knowledge base.

In addition, to test certain community-level programs (e.g. Communities that Care) one would need a large sample of communities to have sufficient power to detect differences. This can offer substantial problems for randomized designs and here again quasi-experimental designs involving comparison communities make sense.

In Pennsylvania, we have been evaluating the first twenty-one communities involved in Communities that Care across Pennsylvania. We are using a very complicated methodology called propensity analysis to match the twenty-one counties with thirty-seven other counties that are not involved in CTC and seek to draw conclusions. We have received some evidence that CTC counties have experienced greater reductions in delinquency rates than comparison communities, factoring in a number of contextual variables related to the counties (Osgood et al., 2001).

We should employ programs that have shown an impact through RCTs where possible. If they are not employed, we should require a strong justification for *why* they are not appropriate.

Another alternative conceptual perspective has often been called comprehensive community action research (CCAR; Weissberg & Greenberg, 1998b). This involves the active involvement of those developing and implementing new approaches in designing evaluation strategies, with the evaluator helping to refine assumptions and hypotheses and devise ways of testing them, as the work proceeds. CCAR questions how much simply importing programs from other locations (even with strong clinical trials) will work in diverse communities.

CCAR argues that one key to program success is that people in the community have strong ownership of and belief in the program. Therefore, this model champions the notion that it is important to empower local solutions that fit the culture of the community. A second notion is that the solutions that should be employed should carefully utilize “best practices.” At the same time, however, there needs to be rigor in program design and evaluation that can test for desired impact and that can begin to identify core activities that are essential to success. This research would ideally combine both quantitative and qualitative approaches. Finally, there is a need to learn which aspects of the program may be generalizable across communities. Again, it is desirable that some programs or some program elements can be further tested by randomized controlled trials.

I have several conclusions regarding the use of evaluation research in funding decisions.

First, we should recognize that one size does not fit all, and we need an overall approach of “planned critical multiplism” (Shadish, 1986, 1990). From the beginning, we ought to think about the different “rich clues” we can collect to help answer the important questions we have. We need to use all these clues and tools. This involves randomized trials and quasi-experimental designs. It also involves qualitative data, interviews with staff, interviews with families, and assessments of the community and its culture.

Second, we should employ programs that have shown an impact through RCTs where possible. If they are not employed, we should require a strong justification for *why* they are not appropriate.

Third, in program areas with little or no RCT data, such as after school programming, we should rely upon best practices and move toward quasi-experimental or RCT designs.

Fourth, we need to support qualitative as well as quantitative research. Qualitative research is essential in helping to understand why a given program has produced a positive result. An RCT can tell us whether a program produces a result, but it cannot tell us what we need to know about why it works.

Finally, we need to recognize that all programs are likely to require tailoring to local circumstances. We need to be explicit in

As with research on effective programs, we need research on effective implementation, a research field that itself is still being built.

how and why we are tailoring those programs, and we need to be concerted in determining what impacts this tailoring may have had.

This final point leads to the second part of my presentation, which involves ensuring quality in program implementation. When a community selects an empirically validated program for funding, it has only started the process of achieving the results that program promises. It must maintain high fidelity with the program model, and it must understand what factors influence implementation quality, as well as where adaptations are possible and where they will violate program integrity. Selection is only the first step in the process; the manner the program is implemented is critical to its success (Domitrovich & Greenberg, 2000).

As with research on effective programs, we need research on effective implementation, a research field that itself is still being built. We do know some things about effective implementation – factors that relate to pre-planning, supports for implementation, the implementation environment, and support for implementers — that should be incorporated into our implementation strategy (Greenberg et al., 2000).

If adequate pre-planning for program implementation is not conducted, it is likely to dramatically affect program implementation. Laying the groundwork is critical and requires several steps and activities. First,

it requires developing community awareness. Second, it requires developing community buy-in. Third, it requires creating incentives to change in the institutions or systems that will conduct or support this innovation. Fourth, it requires dealing with past histories of new program implementation.

Effective programs do not operate in isolation. They must be known and valued in the community — by participants, by referring sources, and by other public and private systems. Their staff must feel valued and recognized. Many communities have had experience with programs “parachuted” in and seen as threats or competition to other systems. This makes it difficult, if not impossible, for those charged with implementing the program to build the relationships that are needed to make it work and lead to sustainability.

Most communities simply do not change very easily. We need to do the pre-planning to support change, including the incorporation of new programs. Working in Pennsylvania with Communities that Care and Family Service System Reform (FSSR), we know these processes are slow and we need to create incentives for people to make and adapt to change. Further, given a past history of failures in implementation, if a new program is to be successfully implemented, it will have to distinguish itself from those prior, failed efforts. Implementation support systems are critical. Alternatively, there may already be a program in

There are multiple dimensions to implementation that require understanding if we are to use research-based programs effectively in our communities.

place with very similar goals and a constituency of people who support it. Pre-planning must address how to address, or at least prepare for, potential conflicts.

Support for program implementation. In addition to pre-planning at the community level, it is necessary to provide adequate support for program implementation itself. This includes adequate staff orientation and training, ongoing technical assistance and support through the initial development stages, and ongoing communication and problem solving throughout the program operation. This is much more than a two-day training program that describes program elements and features. Most research-based programs require highly skilled staff that are able to respond to a variety of presenting situations. Training and support for the development of this staff goes beyond short-term training and requires an infrastructure that supports ongoing development and response to issues as they arise throughout program implementation. This requires effective communication and support from those with experience operating effective programs.

Leadership in the implementation environment. Third, there needs to be attention to the implementation environment and its mission and goals. This includes the critical factor of developing administrative leadership. Depending upon the program, this may require leadership from city and county governments, the schools, and human service administration. It is important that the

program have connections with other service providers and systems and that there is a positive organizational climate. In addition, such leadership can help ensure that sufficient resources are allocated to support the program, and that there is ongoing support to sustain and, over time, institutionalize the program.

Support for implementers. Fourth, there must be support for those on the ground who will be implementing the program. This involves the training and support described above, but it also includes ongoing contact to insure that implementers feel sufficiently prepared and includes monitoring to insure that implementers are not overstressed or feel neglected and unappreciated. It is often very stressful to complete all the tasks necessary to get a program up and running, and this hard work needs to be recognized and supported.

In summary, there are multiple dimensions to implementation that require understanding if we are to use research-based programs effectively in our communities. We have to have a strategy in place to assess and use our implementation experiences. We have to develop tools to understand why a program worked better in Jefferson County than in Erie County. We need to learn what strategies are successful in involving local government, school districts, and community agencies in planning and implementation and what impacts this involvement produces. We need to learn how our training systems may be affected by

We need to gather lessons learned from our implementation experiences so subsequent efforts in other communities can build upon those experiences.

local factors and conditions. We need to gather lessons learned from our implementation experiences so subsequent efforts in other communities can build upon those experiences. What one community experiences in implementing a research-based program should be information for use by the next community in its own implementation.

In 1977, Hubert Humphrey said that “the moral test of government is how it treats those at the dawn of life – children.” In most states, including Pennsylvania, we have not paid sufficient attention to this issue and to how children are embedded in families, schools, neighborhoods, churches, and communities. The movement in the last eight years in Pennsylvania toward more collaborative responses is a critical element of a positive response to this moral test of government. It is not just about providing services to children; it is about thinking about ways to get communities to come together for children and to use effective programs and strategies to get better results. This task will be more effectively accomplished with greater accountability as we better utilize multiple methods of research and evaluation to ensure our public dollars are truly spent on effective programs.

References

Domitrovich, C., & Greenberg, M. T. (2000). The study of implementation: Current

finding from effective programs for school-aged children. *Journal of Educational and Psychological Consultation*, 11, 193-221.
Dryfoos, J.G. (1998). *Safe passages: Making it through adolescence in a risky society*. New York: Oxford University Press.

Ennett, S. T., Tobler, N. S., Ringwalt, C. L., & Flewelling, R. L. (1994). How effective is drug abuse resistance education? A meta-analysis of project DARE outcome evaluations. *American Journal of Public Health*, 84, 1394-1401.

Greenberg, M. T., Feinberg, M. E., Osgood, D. W. (2001). *Pennsylvania CTC Evaluation Update*. CTC Technical Report #4, Prevention Research Center, Pennsylvania State University.

Greenberg, M. T., Domitrovich, C., Graczyk, P. A. & Zins, J. (2000). *The study of implementation in school-based prevention research: Theory, research and practice*. Report to the Center for Mental Health Services (SAMHSA); Washington, DC.

Lantz, P. M., Jacobson, P. D., Warner, K E., Wasserman, J., Pollock, H. A., Berson, J., & Ahlstrom, A. (2000). Investing in youth tobacco control: A review of smoking prevention and control strategies. *Tobacco Control*, 9, 47-63.

- Mrazek, P. J., & Haggerty, R. J. (Eds.). (1994). *Reducing risks for mental disorders: Frontiers for prevention intervention research*. Washington, DC: National Academy Press.
- Olds, D. L., Hill, P. L., Mihalic, S. F., & O'Brien, R. A. (1998). *Prenatal and infancy home visitation by nurses*. Blueprint for Violence Prevention (Book 7), Institute of Behavioral Sciences, University of Colorado.
- Osgood, D. W., Anderson, A. L., Feinberg, M. E., Greenberg, M. T. (2001). *County level data analysis of CTC impact*. CTC Technical Report #5, Prevention Research Center, Pennsylvania State University.
- Powell, D. R. (1994). Evaluating family support programs: Are we making progress? In Kagan, S. L. Kagan and B. Weissbourd, (Eds). *Putting families first: America's family support movement and the challenge of change*. (pp. 441-470). San Francisco, CA: Jossey-Bass Inc.
- Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2001) Long-term effects of an early childhood intervention on educational achievement and juvenile arrest: A 15-year follow-up of low-income children in public schools. *Journal of the American Medical Association*, 285, 2339-2346.
- Schorr, L. B., & Yankelovich, D. (2000). What works to better society can't be easily measured. *Los Angeles Times*, February 6.
- Schorr, L. B. (1987). *Common purpose: Strengthening families and neighborhoods to rebuild America*. New York; Doubleday.
- Shadish, W. R. (1986). Planned critical multiplism: Some elaboration. *Behavioral Assessment*, 8, 75-103.
- Shadish, Jr., W. R. (1990) Defining excellence criteria in community research. In P. H. Tolan, C. Keys, F. Chertok, & L. Jason (Eds.), *Researching community psychology: Issues of theories and methods* (pp. 9-20). Washington, DC: American Psychological Association.
- St. Pierre, R. G., Layzer, J. I., Goodson, B.D., & Bernstein, L. S. (1997). *National impact evaluation of the Comprehensive Child Development Program*. US. Department of Health and Human Services, Administration on Children Youth and Families: Washington, D.C.
- Weissberg, R. P., & Greenberg, M. T. (1998a). School and community competence-enhancement and prevention programs. In W. Damon (Series Ed.) & I. E. Sigel & K. A. Renninger (Vol. Eds.), *Handbook of child psychology: Vol 4. Child psychology in practice* (5th ed., pp. 877-954). New York: John Wiley & Sons.
- Weissberg, R. P., & Greenberg, M. T., (1998b). Prevention science and community collaborative action research: Combining the best from both perspectives. *Journal of Mental Health*, 7, 477-490.

CHAPTER 3

Research and Policy: A Cross-National Perspective

Michael Little

As you can tell from my accent, I am English and my modest role on this panel is to provide some cross-national perspective on research and policy.

My professional work has been in linking research and policy, in drawing from the best evidence available about what works to help develop child welfare policy, in particular, in Great Britain. I am currently in the United States to learn more about your work and research. Some of the best scientific research on children's issues is done here in the United States. At the same time, however, I suspect that we in England do better than the United States in integrating research into public policy.

One aspect of English people you should know about is our pragmatism. We are not for or against anything. We generally believe there is a place for most things in all formulations. I am often amazed here in the United States that people come up and ask, "Are you for or against adoption?" or "Do you think our family preservation services should be scrapped?" or "Are you an advocate or critic of randomized control trials?" To an English person, these do not seem to be very sensible questions. It is a little like asking whether one is for or against hospitals. I am for it in the context of necessary surgical intervention; I am against it in the context of a headache.

In the prevention world, one sometimes hears the question, "Is prevention better than intervention?" If you look at most

examples of strategies to improve child welfare that have worked around the world, it is the combination of prevention with early intervention with intervention with social prevention, along with much better diagnostic techniques that leads to improvement. When improvement is the goal, it is all those things in combination that produce a difference, not any one factor working alone. No successful strategy that I know of has ever totally eradicated a social problem. Yet we tend to talk about trying to end anti-social behavior or trying to end violence in schools.

In short, we need to look much more in terms of degrees of success, and not in "all or nothing" terms. We should not simply try to judge whether a program works or does not work, but we should look at what level of impacts can be expected under what conditions, with what specific populations and in combination with what other services.

That brings me to a fairly simple idea about where to start with our work. I would not start at the program level. On the whole, I think it is best to start our inquiry with evidence about children living in our community. What do we know about child development or the children living in our communities?

My work in England has focused on identifying those children who have serious problems, what in England we call children in need. We try to get all of our agencies —

We need to look much more in terms of degrees of success, and not in “all or nothing” terms.

health, education, social services, etc. — to come together to establish a common definition of children in need, defined as those whose development is, or is likely to be, impaired.

Next, we encourage people to think about outcomes. What is it that we can realistically achieve on behalf of those children in our communities? That is not to eradicate those problems, but to show realistic improvements in child development and well-being.

At this point, we encourage people to think about the services, not necessarily the services currently in place, but services that might realistically achieve specified outcomes. Finally, we try to encourage people to establish an appropriate organizational structure to deliver those services. So that is my rather simple idea, emerging from our rather pragmatic nation.

To a sophisticated nation, such as yours, this simple approach may not seem like a great deal. But when I look at the United States, I often see people starting with an organizational structure into which services are forced and adapted to the financing that is available. Then, the services are delivered, whether or not they are a good fit with children’s needs. At the same time, there is not a common definition about which children have serious needs or what those needs are. Different systems define them very differently. At the end of all this, there can be evaluations and disappointment with

the observed outcomes.

So my rather simple idea is to turn that strategy on its head and to start with the children instead of starting with the organizational structures that currently serve the children. A lot of my work as a social policy researcher, in the last fifteen years, has been to try to turn that formulation around. In England, we have had some success, including legislation enacted in 1990 that requires each community to bring all its agencies together to work with consumers to know which children have social needs and then to use that information to design services to meet those needs. We have not worked out all the details, but we are making progress. We have started to break down the different languages and terminology across different fields and to develop more holistic and effective responses.

I want to say a little more about the issue of a common language. If I say juvenile delinquents in the United States, many people immediately think of violence and gang behavior. In England, it is more likely to be children stealing from shops or cars. If I say child protection, in a room of one hundred people, we would probably get fifty different answers. The word prevention would probably produce five hundred different answers, five per person. These are words that can mean very different things in different contexts.

Winston Churchill has said the United States and the United Kingdom are two

The lack of innovation is a huge handicap in family and children's services. We continue to do most of our work the same way it has always been done.

nations divided by a common language. I think that also is true also true of practitioners, policy makers, and researchers. Very often, I see two practitioners sitting across a desk from one another, looking at the same information of a particular child and coming up with very different diagnoses and courses of action. We don't even have one language when we are working with the same child, which confuses our interventions, and us and reduces the impact we have on the child and the child's family.

Some of my recent work has involved working with people in ten communities in England, Norway, Spain, Italy, Ireland and here in the U.S., to see if we can achieve a common language to describe children in need and use that language to learn more about the best aspects of services in each of the target countries. The goal is to avoid endlessly re-inventing the wheel and joining forces to further our knowledge of what works.

It is not an easy task, as you can imagine, but we are helped by one unifying factor across the ten sites. In all of the ten countries, the unifying factor is the children. The children's needs are the same everywhere we have looked. The pattern of need is constant, although the volume differs by virtue of the context in which children are raised. But, their needs are the same. By developing a common language, we may be able to compare our different approaches and be more creative in our responses.

As Ms. Schorr has suggested, we are not approaching our work with a bag full of answers, but we are approaching each of the sites with a bag full of tools to encourage each site to think differently in terms of children, and to be innovative about how best to respond to children's problems. The lack of innovation is a huge handicap in family and children's services. We continue to do most of our work the same way it has always been done in the past, according to the same organizational confines, without clear evidence about what we are actually achieving.

I want to provide an example of creative thinking that you may find worth exploring. All of us on the panel have been involved in working with discreet groups of children, developing new services, and using different evaluative methods to determine whether or not they work. Generally these discreet groups tend to be at the extremes. An alternative argument is that we should change normative behavior in order to affect those at the extremes. Gordon Rose wrote a book about preventive medicine, and made the following observation, looking at the distribution of people's drinking habits.

The United Kingdom has a higher rate of drinking among the general population than the United States, and also a higher rate of alcoholism. Instead of trying to develop new treatments for alcoholism, Rose argued, it might produce a greater overall impact to work to reduce the amount of alcohol that normal people in society drink,

America's children have many serious needs that you must try to address, whether or not you have all the answers on how to address them.

which would move the overall drinking curve and therefore reduce the number of alcoholics. This is a different way to look at the problem. I offer this example to encourage some reflection on our orthodox approaches.

In closing, I would like to commend your work here in Pennsylvania. It is very important to draw from research and to get a better understanding of what it takes to improve child development and well-being. It is especially important in the United States. There are many things I see here in the United States that I am pleased we do not have in the same degree in England. We do not have your great poverty or the huge gaps between rich and poor. We also do not have your level of poor child outcomes, in health, education, and social adjustment. We have children in need, but we do not have as many as you and we generally provide more supports to children than you do.

As you move forward, I dare you to be brave, or perhaps fool-hearted, as we in England have been. Without much support for the idea, we introduced universal health visitor services for all newborn babies in the 1940's. I was relieved, when my wife and I had kids, to see that home visitor at our door on day one to tell us what we needed to do. All children in England are covered, and if you brought your children to England they, too, would be covered. These things have been introduced without much evidence that they would work, but they have made huge

differences in child development. If they had not worked, I believe we would have abandoned them, but we took action because we started with an identified need.

America was brave enough to build a rocket and put it on the moon, I think you need to be much braver when it comes to children's services. America's children have many serious needs that you must try to address, whether or not you have all the answers on how to address them.

In America, it always seems that we are talking about "other people's children," whereas in England, we tend to refer to all of our children as "our children." It may be that this simple cultural difference has an impact upon how we respond to children's development.

CHAPTER 4

Making Progress: Learning and Public Accountability

Heather Weiss

I believe this seminar and this topic are very timely. For years, I personally have been puzzled about how to get research and evaluation-based information into play to help providers deliver effective and accountable prevention services. I believe that the stars are now aligned to make this happen in a way that it has not in the past. Over the next five to ten years in this country, I believe it will be possible to link research and evaluation-based information with practice, to improve outcomes for children and families in a continuing way on a widespread basis. I want to talk about how we can make that happen.

First, I want to discuss why this is possible. Then, I want to talk about a framework that needs to be put in place at the program, organizational, state and national levels for it to happen. It will be evident as I proceed that the work of the other people on this panel is critical to conducting quality research and distilling the lessons learned from that research. What I want to talk about is how we are going to use that information to improve outcomes for kids. At this point, I think we invest more in gathering information and doing research than we do in putting that information into practice to support the improvement of services to children and families. I am going to assume that we will continue to gather information and conduct research. I want to talk about how we are going to use that information so that providers can be accountable and have an opportunity to continually improve what they do.

I am going to argue that the earmark of a quality program or organization is that it has the capacity to get and use information for continuous improvement and accountability. No program, no matter what it does, is a good program unless it is getting and using data of a variety of sorts, from a variety of places, and in an ongoing way, to see if there are ways it can do better.

Why are we poised to link research and evaluation-based information with practice, to improve outcomes for children and families in a continuing way on a widespread basis?

It is because there have been big changes in the public policy and legislative context at the state and federal levels that are creating new demands for research-based information. In the past, once a program model had been developed and had research attesting to its effectiveness, program effectiveness was largely assumed and accountability was based upon fiscal compliance and demonstration that funding was expended as intended for the program.

In 1990, the federal government passed the Government Performance Results Act, and with that a new age of accountability began. Governments began to call for evidence of program impact, and with that came a growing demand for research-based information. In turn, this has created a growing demand for help in getting and using data to manage organizations that serve children and families for continuing high perfor-

The earmark of a quality program or organization is that it has the capacity to get and use information for continuous improvement and accountability.

mance.

A good example of how far this has progressed is an e-mail I received two weeks ago from Judy Jerald, who manages the Early Head Start Program within the Head Start Bureau. She was visiting a particular state, and this problem was presented to her: key legislators did not know much about Head Start, but they were not interested in what national studies and reports said. They wanted state-specific information. They did not want more studies, but they wanted evidence that the specific Head Start programs in their state produced results. This is the reality of life out there for programs, even programs like Head Start, which we think of as having secure federal funding. The old way of doing things — develop a model, show its effectiveness, and replicate it in three places and say it works — does not work so well today.

This poses huge challenges for researchers and evaluators, who can no longer argue that a particular program works on the basis of a national study. Nearly everyone needs evidence that their specific program will produce results to make the case for new funding.

I have been spending a lot of time over the last year looking at non-profit organizations around the country who are trying to manage the move to outcomes. They are finding a near universal demand for demonstrating outcomes as part of their funding accountability, whether that funding is from the

public sector, from the private sector, or from United Way or a foundation. Funders are all saying, “We want to see evidence that your programs are making progress in achieving outcomes for the children and families you serve.” They have equated their fiscal responsibility of insuring money is well spent with demonstrating improved outcomes.

I think we must recognize that we currently have a lot of ineffective programs and practices out there, as they currently operate. As time goes by and more programs are required to show their results, I think this will become clearer and clearer.

At the same time, there is relatively little investment in communicating information about effective program models and practices and how to implement them. There is little investment in building organizational capacity for evaluating and tracking program performance and program outcomes for use in self-improvement. There is huge pressure to do outcomes and be accountable without supports to do so.

I think we need to transform the accountability discussion. As accountability matures, there will be increased and continuing demand for research-based information about what works and an increased willingness to use that information in funding decisions. If we are not careful, we will be at risk of another cycle of disinvestment in social programs, based upon the conclusion that “nothing works.”

We have to share lessons learned about what has worked and what has not, and we have to adopt a continuous improvement approach that demonstrates that programs learn from their experiences.

This “gotcha” accountability is likely to occur unless we help local providers to get and use data in a way that satisfies funders that they are making progress in achieving outcomes and are using their experiences, and the experiences and research that exist in the field, to improve their practice.

I think there are multiple ways for research and evaluation data to inform policy and practice.

One approach is to replicate and scale up with fidelity a research-based program model and demonstrate positive outcomes in that scaled-up effort. We have heard a little bit about the importance of fidelity in replication and adaptation from other panelists. I think we are also under an obligation to share the information we have obtained about the successes and failures in these scaled-up efforts, to inform others and to build a body of knowledge about how to effectively scale-up research-based programs and, hopefully, adapt programs to meet local needs and improve upon the original design.

A second approach is to start with a general model, hopefully employing attributes of effective practice, that seeks to tackle an important concern, where there may not be an identified, superior research-based program that is appropriate to the situation. States have established funding for general program models set out to improve results in a specific area, such as improving school readiness or family stability or

reducing teen pregnancy. In these instances, we need to test and track performance, with a clear eye on positive outcomes. We have to share lessons learned about what has worked and what has not, and we have to adopt a continuous improvement approach that demonstrates that programs learn from their experiences to improve practices and achieve better results.

In both of these approaches, but particularly the first approach, we have to address what I call the Gina Collada challenge. In preparing for a story for National Substance Abuse Week, Gina Collada, a premier science writer for the *New York Times*, did a little poking around and found out that research pointed to the fact that a much used and well-funded program, Drug Abuse Resistance Education (D.A.R.E.) has evaluations indicating it does not appear to reduce drug use among students. That is not a major revelation; this is fairly well known in the research community. At the same time, D.A.R.E. is in 99% of the school districts in the country, so we have a delivery system that works, although it is implementing a probably ineffective program, at least with respect to reduced drug use.

She did not stop here, however. She then found the federal government had invested a lot of money in a university-based researcher who had developed a school-based substance abuse prevention program that worked. It was unequivocal that it worked. It had been replicated in several sites, with

We need to develop a commitment to learning — to continuous improvement, accountability and program organization.

results showing that it worked under several different conditions with several different populations. This program never was more broadly disseminated, however. She asked the researcher who developed the program why, and he said it was his job to do research and not to develop implementation strategies for program replication expansion. She then called the federal officials who funded the research, and they said that they did not have any apparatus to communicate the research findings or develop an implementation strategy for expansion. Therefore, a strong, research-based program sits on the shelf, even though there is a delivery system like D.A.R.E. that might make very good use of it. This is the Collada challenge. This does not let D.A.R.E. off the hook in terms of continuing to fund ineffective programming, but it does let them off the hook in being fully responsible for developing a more effective strategy. We need to develop some responsibility and capacity for assuring that relevant research and information gets out about what works to the people who most need it. I think we have got to think about how we can successfully address the Collada challenge.

I do not think that there is a simple answer, such as a new program office for the diffusion of effective programs. I think the challenge is deeper than that.

I think we need to develop a commitment to learning — to continuous improvement, accountability and program organization.

Learning is at the core of this. David Garvin of the Harvard Business School defines a learning organization as one skilled at “creating, acquiring and transferring knowledge and at modifying its behavior to reflect new knowledge and insights.”

We are beyond the point where we find a proven program, get it in place, and then put it on “auto-pilot” and say it worked in other places so it works here. We are in a different ballgame than we have been in before. More than ever, we need good information about what works and we need to figure out how to get that information into the hands of people that can use it, and then provide them with tools and support to assess if and how it works in their particular place.

We are setting the bar very high in the kind of game that we are in now. It involves creating a culture of learning within organizations and programs. Providers have to identify outcomes and measurable indicators of those outcomes, even if those indicators are not perfect, but that is only the start. Providers need to continuously measure what is happening, collect and analyze the data on these indicators and other factors, use this data and analysis to make program changes and adjustments, while continuously reporting on what has been done, what has and has not been achieved, and what is being changed as a result of this whole process.

I want to give you several examples of this

We badly need more partnerships between community providers and university-based researchers – especially genuine partnerships where learning is going both ways.

approach.

Two are drawn from a national study we are conducting for the Pew Charitable Trusts. It is a set of case studies of non-profit organizations that are making this transition to be learning organizations and managing for high performance.

One of these is a youth development organization on the West Coast that is based in a research-based conceptual framework about what adolescents need for development. One element in that framework is a safe and secure environment. The organization did a survey of the kids at their out-of-school-time program and found out that the bulk of them did not feel safe and secure. They had to modify a whole set of their own practices so that participants would feel safe and secure in their program. They got data from participants, they analyzed it, and they made changes to the program to give them a better chance of achieving their long-term outcomes.

The second of these is a program that provides services to women making the transition from welfare to work. The program wanted to add an additional service that they believed would increase program success. They found a funder who was willing to fund that service, but only if the program studied its impact through a randomized design, where the service was provided to half the population being served. They tested the service over a six-month period, and found very compelling

information that the additional service did produce gains. They now can make the case for adding the additional service based on this mini-experiment.

A third example of organizational learning and change is drawn from an Early Head Start (EHS) meeting where I recently discussed some of these ideas. One of the EHS local program directors noted that after she got the results from the national cross-site evaluation of the Comprehensive Child Development Program (CCDP) in which her program participated, she and her staff did some real soul-searching. They assessed why they had failed to produce gains and what they needed to do differently under their new EHS funding. Participation in the EHS evaluation enabled them to pair up with a university-based researcher. Together, they carried out new research that identified some things about their population that enabled them to better tailor their services to meet their needs and to correct the problems they had identified in their implementation of CCDP.

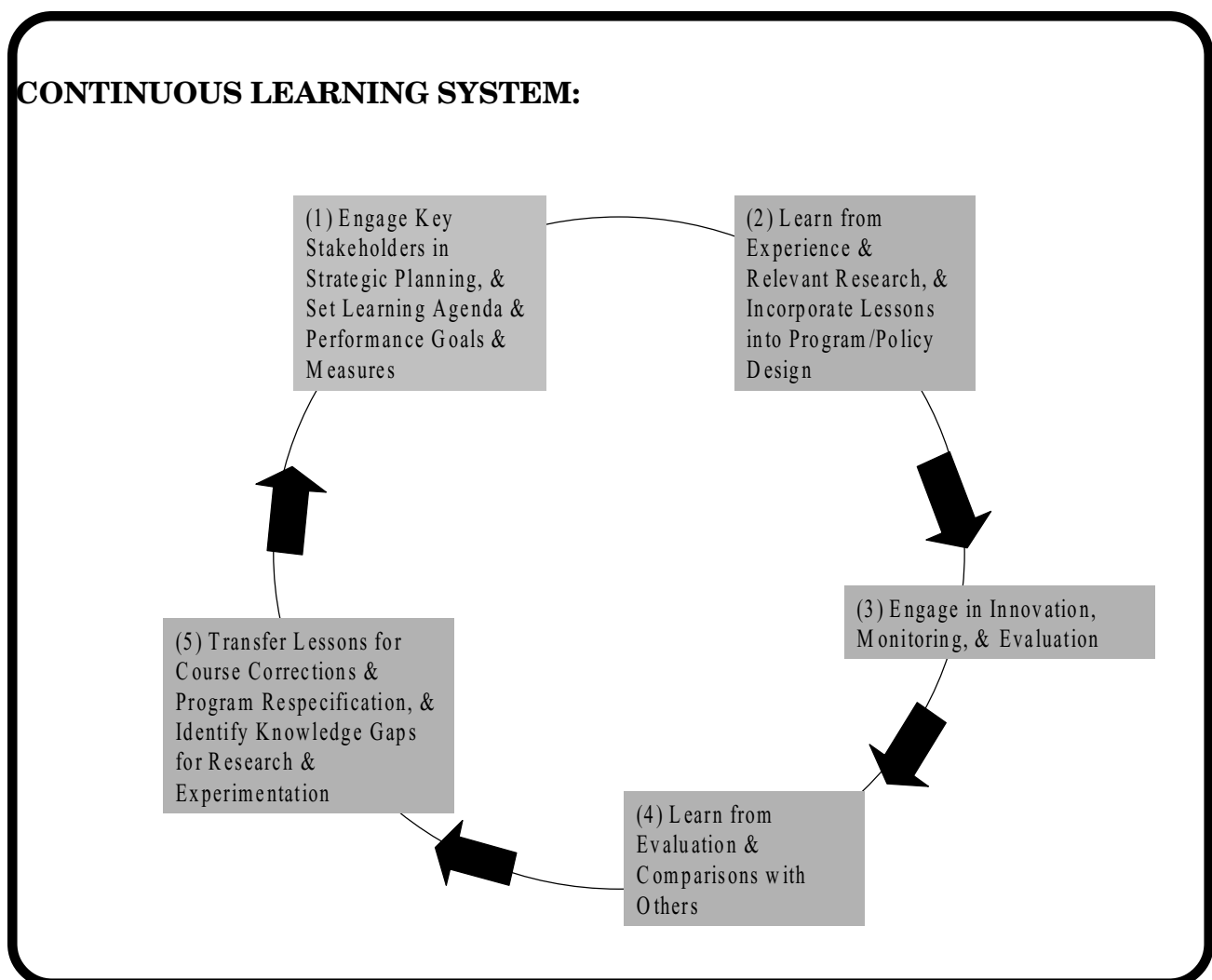
This third example raises a very important point. We badly need more partnerships between community providers and university-based researchers – especially genuine partnerships where learning is going both ways. That should be an earmark of the way you invest here in Pennsylvania, because building those partnerships is essential both to learning and to the use of that learning in improving practice. If we are going to have learning organizations, we

We are obligated to find out what is already known about what works and does not work and incorporate it into our design.

must support them in their performance management efforts with good tools and research, including evaluation.

Some of the kinds of support that are needed are displayed in the Chart below.

Walking through this variant of a learning system, it begins with the view that when we design new programs or strategies, we are obligated to find out what is already known about what works and does not work and bring that information in from the outside and incorporate it into our design.



If we are going to expect programs to meet this higher standard of learning, accountability and effectiveness - they are going to have to have the financial, organizational and technical resources to do so.

This means we have to establish ways to make that information accessible on an as-needed basis, which means we have to invest in the necessary infrastructure. We have to collect and synthesize research as my fellow panelist, Mark Greenberg and others do. They carefully address the questions of what do we know about proven programs — and make the answers accessible to programs and organizations that develop and provide services.

Oregon has developed such an information infrastructure for their state work on children and families. They track relevant research and evaluation in an ongoing way and post it on a website that is periodically updated with new evidence and experience.

The next step on the learning system graphic positions program performance, monitoring and evaluation to provide feedback in implementation, program and participant performance and outcomes. The assumption is that programs will use this information for course corrections and learning. This also includes a benchmarking process that enables comparisons both within the program and with other programs working toward similar goals. While programs need to do self-evaluation, they also need external supports and resources for comparisons and benchmarking with other programs. Therefore, we need learning organizations embedded within a larger learning system that provides information about effective programs and practices, about appropriate

benchmarks and about relevant basic and applied research.

The next step in the learning process is to transfer the lessons learned from program efforts for program re-specification and to identify knowledge gaps for further research and experimentation. By clearly identifying what are true gaps in our knowledge, we can support experiments that are not re-inventing the wheel but truly expanding our knowledge base. This is experimentation nested in what we need in the field to make progress toward better results. It is the strategic use of experiments to fill the key knowledge gaps that we have identified in practice. This is a very high standard for learning and accountability, as it involves not only determining whether programs are effective but insuring that we are moving to make them and our overall social problem solving more effective.

This argument for learning organizations within a larger learning system has important resource allocation implications. If we are going to expect programs to meet this higher standard of learning, accountability and effectiveness - whereby they can get and use data for learning, continuous improvement and accountability - they are going to have to have the financial, organizational and technical resources to do so. At present, programs are fortunate if they have any money for this purpose. Our research in non-profits making this transition to becoming learning organizations

Unless we help social programs make the transition programs are not going to be prepared and many will be victimized by loss of funding.

suggests that programs or organizations must have sufficient and sustained resources and that they are very hard to get. Funders increasingly want outcome-driven programs but rarely provide the necessary upfront and then continuing resources.

This argument also means that we must build an information infrastructure to provide the necessary outside support, information, and technical assistance to learning organizations. Within a state, this probably requires a consortium, a public and private consortium, to provide information about what works and to provide support in developing outcome-based organizational self-evaluation and continuous improvement systems. This requires partners, which include universities and researchers, who can work with organizations in practical, real-world ways, while contributing expertise in research and evaluation design.

This is a huge challenge. If we do not accept it, I think we are rapidly going to be in a situation where many funders, both government and private sector, conclude, “We do not know what works, we are not making any progress on trying to figure it out, and we are not going to fund it any more.” This can lead to massive disinvestment for social problem solving in the community.

If you take away nothing else from this seminar, talk to your neighbors who run agencies, talk to your neighbors who sit on United Way boards, and talk to your elected officials who fund programs. They will tell

you there is hardly a funding source that is not focusing on outcomes and asking social service programs, in particular, to justify their funding by stating what outcomes they will improve. Two years from now, if not sooner, these funders are going to want to see evidence that these programs have made progress on achieving these outcomes. Unless we help social programs make the transition to the framework for outcomes and accountability and learning I have described, programs are not going to be prepared to respond and many will be victimized by loss of funding. We are at risk of massive disinvestment in social services and prevention efforts that we ultimately need — done better and more effectively — for our country’s children and families.

You are poised here in Pennsylvania to help your state programs for children and families make this critical transition to learning and accountability. If you can do so, you will be in the vanguard of those making real, measurable progress in improving the lives of families and the life chances of their children. You will also, by your example, help many other states and communities struggling with similar issues. If work here in Pennsylvania and elsewhere in the country really dedicates itself to this task and this seminar and Pennsylvania’s actions show promise in this direction, we can have more effective programs and services that make progress toward the results we all desire for our children and families. Thank you.

CHAPTER 5

Developing Outcome-Based Initiatives: A Foundation Perspective

Cynthia Guy

I am here to provide a Foundation perspective on the use of research to design effective programs, based on the work of the Annie E. Casey Foundation. Casey is a private philanthropy whose central mission is to improve outcomes for children and families. The Foundation has an interest in research as both a consumer and producer of research. We are consumers of program evaluation in our efforts to design and run effective programs based on the best evidence we can bring to bear. We also produce research as a funder of evaluations that are designed specifically to expand the knowledge base around what works. We do this both for our own benefit as program designers, and in an effort to influence public policy and practice more broadly.

As a consumer and producer of evaluation, the Casey Foundation has developed a fairly broad interpretation of what constitutes research evidence. For example, the Foundation frequently supports and promotes the most rigorous forms of evaluation for programs, including randomized controlled trials for programs that meet the necessary conditions. Such programs include those that deliver very specific, quantifiable treatments on a large enough scale to support statistical research.

At the same time, the Foundation also acts as a designer and implementer of programs, and in this role, we are keenly aware of the limits of randomized controlled trials in guiding program development. I want to emphasize a couple of those limiting factors,

which influence how the Casey Foundation uses research to develop programs. In doing this, I will draw on the example of the Plain Talk Initiative, which was designed and operated by the Foundation, and the Plain Talk evaluation, which was conducted by Public/Private Ventures (P/PV).

One factor to consider is that the impact of tested program models is often so unsatisfactory that there is an obvious need to go beyond those models, to do more than replicate them. In some fields, the evidence produced by rigorous evaluation seems to tell you more about what does not work than about what does. In such cases, we need to learn how to use research not to replicate, but to innovate. When you are innovating, you don't have validated models to guide you, but you can draw on other kinds of data and research to help you improve upon approaches that have been tested and proven inadequate.

Let me give you an example of how the Casey Foundation went beyond replication of experimentally validated findings. Casey's Plain Talk demonstration was designed to prevent adolescent pregnancy and disease among sexually active teens. The strategies for achieving this goal focused on making reproductive health services more available in local communities, improving communication between adults and youth about sex and contraception, and getting youth to use protection and avoid risky behavior.

Good programs, programs worthy of replication, may not allow for experimental validation.

The Casey Foundation sponsored Plain Talk in target neighborhoods within five cities from 1993 to 1997. Plain Talk was not a replication of a research-based program model. It was, however, rooted in research and evidence of a different kind. To a large degree, the design of Plain Talk was informed by the negative findings of previous evaluations of teen pregnancy prevention efforts. These included the disappointing results of programs that focused on providing pregnancy prevention services to individual children, without working with or through their families or communities. Beyond this negative evidence, program designers were guided by reviews of the research literature comparing the United States and Europe in terms of sexual activity rates amongst teens (which are very similar) and teen pregnancy rates (which are much higher in the United States). In looking for an explanation, this research focused on the differences in the way Europeans and Americans deal with teen sexuality. In particular, it found a much greater willingness among adults in Europe to discuss sexuality frankly with their kids and their proactive support of using protection. The Casey Foundation also conducted its own domestic data gathering efforts. We did literature reviews of American research on teen pregnancy prevention programs, and we did focus groups and in-depth discussions with researchers and knowledgeable practitioners in the teen pregnancy field that were designed to identify the gaps in United States pregnancy prevention programs and to suggest how they might be

filled.

The evidence, the data, and the lessons gathered in the program design phase led the Foundation to develop an initiative that involved local adults from the family and community in getting sexually active youth to protect themselves from pregnancy and disease.

The Plain Talk design process was a conscientious effort to take research findings – both positive and negative – and use them to inform the development of a new service strategy for a social problem that badly needed one. This process was a far cry from replication of an experimentally validated model, but it had to be – there was no tested model out there that could do the job we needed to do.

Turn now to a second factor that argues against an exclusive use of experimental research to inform program development. The second factor is that good programs, programs worthy of replication, may not allow for experimental validation. In fact, sometimes it is the very characteristics that make them successful that also make them impossible to evaluate by these means. We are talking here about small-scale programs, where you cannot get sufficient sample sizes for statistical analysis necessary for impact studies. We are talking about saturation programs, where eligibility is defined by residence in a particular geographical area, and by definition, everybody in that area is eligible for services. We

We are talking about programs in which context is not a variable that can be controlled for, but is in fact integral to the treatment.

are talking about programs staffed by volunteers or community activists, people that are not going to be very amenable to enforcing the kind of discipline that random assignment requires, and are in fact likely to go out of their way to provide services to the control group, if random assignment is attempted. Most importantly, we are talking about programs in which context is not a variable that can be controlled for, but is in fact integral to the treatment. In this category are programs designed to draw on community assets as part of the intervention, and programs designed to actively adapt and evolve in the course of the intervention.

At the Casey Foundation, we have been forced to appreciate these barriers to experimental design evaluation. Many of the programs we develop are saturation programs whose main goal is community-level change and whose primary strategy is resident involvement. We cannot do randomized controlled trials, but we have not given up on the goal of establishing other kinds of counterfactuals that we can use to test whether or not we are having an effect.

We find an example of this again in the Plain Talk Initiative. In designing and implementing Plain Talk, we wanted to contribute to the knowledge base about teen pregnancy and sexually transmitted disease. Therefore, we wanted to put the strongest evaluation we could into place. The Foundation recognized from the start that we could not randomly assign teens to

treatment and control groups in our neighborhoods. Random assignment would have been confounded by a key element of program design – Plain Talk is a saturation project, intended to reach teens indirectly through a community-wide effort that changes the whole ethos of a community, particularly attitudes and messages relating to sex and contraception. In addition, Plain Talk was designed to be run largely by community activists, groups that are ill-suited to enforce research discipline that is inconsistent with their mission. For all these reasons, the Foundation never seriously considered doing an experimental evaluation of Plain Talk.

The Foundation did, however, consider comparison group evaluation, which would have involved comparing outcomes in the Plain Talk neighborhoods to outcomes in a matched set of non-program neighborhoods. It did not take long, though, for both the Foundation and the P/PV evaluators to recognize the unfeasibility of coming up with acceptable matches for the sites selected for Plain Talk. Each Plain Talk site was selected for participation in the demonstration based on a combination of capacities and needs that gave the Foundation some hope that the local actors could fully implement and benefit from this controversial and difficult model. Critical to site selection was our assessment of the leadership and vision of key people within the community. The selection criteria were so complex, based upon both objective characteristics and subjective judgements, that it

We ultimately ended up with an evaluation design that included intensive multi-method research in the program sites.

was difficult enough to identify a small pool of program sites; identifying a pool of comparison sites would have been prohibitively labor-intensive. Nor did we want to scale back on the selection criteria in order to build up a pool of comparison sites. We knew too much about each program site's unique qualifications to be satisfied with a superficial match. In short, efforts to find a comparison community for the Plain Talk evaluation were thwarted by the fact that, in this initiative, the community is not simply the context, it is a partner in the treatment.

We ultimately ended up with an evaluation design that included intensive multi-method research in the program sites. The evaluation conducted by P/PV focused on key Plain Talk strategies and outcomes. The Plain Talk research sites participated in baseline and follow-up surveys, ethnographic research, site visits and document reviews. In order to assess program effects in the absence of comparison or control groups, the evaluators used statistical models to track the relationship between Plain Talk inputs, and youth outcomes. Finally, the researchers did their best to overcome differences in the data to compare national trends in pregnancy and disease rates to trends in the Plain Talk neighborhoods.

In Plain Talk, we feel we have an example of how a broad approach to research can really pay off, in both program design and evaluation. For example, from the Plain Talk

evaluation, we learned important lessons about how to conduct community outreach and education; the data demonstrate that resident-led outreach and education activities reached many more community people than professionally staffed efforts. The evaluation also documented an increase in the availability and accessibility of reproductive health services in the Plain Talk neighborhoods, thus showing that Plain Talk's community-based approach to improving health services is both realistic and effective. Finally, the evaluation produced very encouraging findings on the most innovative and controversial dimension of the Plain Talk approach – the theory that you can protect sexually active youth from pregnancy and disease by improving their communication with adults. Analysis of the survey data supports the idea that kids are less likely to get pregnant when trusted adults communicate with them about sex and protection in a way that is more open, more skillful, and better informed.

My purpose here is not to report the findings of these Plain Talk evaluations, some of which are available on the Foundation website (www.aecf.org) and some of which are still in the publication pipeline. My purpose is to illustrate from one foundation's perspective, the broad set of considerations and conditions that need to be taken into account in determining how we can use research and evaluation to develop effective social programs.

Where we have come out is that we recog-

While we are committed to making the maximum use of rigorous research and evaluation, we cannot allow the limits of current evaluation technology to limit our aspirations to develop programs that work.

nize the value of experimental design research, and embrace the opportunity to produce and use it where program characteristics and conditions permit. We also recognize that the state of knowledge in a given field may require program designers to innovate rather than replicate. In these cases, we have to use research to do more than simply guide us in the implementation of tested models. Finally, we recognize that sometimes, by their very design, really good programs cannot be validated by controlled experiments. Such programs require other methods and approaches. Ultimately, while we are committed to making the maximum use of rigorous research and evaluation, we cannot allow the limits of current evaluation technology to limit our aspirations to develop programs that work.

CHAPTER 6

At the Ground Level: State and Community Perspectives

Tom Gamble, A Community View

I think this has been a very productive exchange of perspectives and shows that research is continuing to evolve and we must keep up with this evolution. I am an advocate for much greater use of research in policy development and funding. I also believe that it is unacceptable to continue funding ineffective programs while we withhold support for programs that can produce results. I want to take my time, however, to raise a concern regarding the imposition of outcome accountability from a ground level, community perspective.

My concern is an almost ritualistic approach to outcome evaluation that is going on now at the community level, where every funder wants an outcome evaluation for every program. We are getting a proliferation of extremely low-fidelity evaluations that often take time, energy and money away from providing programs.

Sometimes, programs are good, in and of themselves. When people get money to run a soup kitchen, to run a food pantry, or to provide a recreation programs for kids, their accountability does not necessarily need to be based upon the long-term outcomes they produce. They should be held accountable to serving people, but to place heavy requirements to determine the impact of their service on the long-term well-being of those individuals is silly. It is silly not only for the obvious reasons that these evaluations take time and resources away

from delivering the service, but because they are likely to be such low-fidelity evaluations that they are not going to yield any useful findings, in any event.

I believe that professional evaluators not only need to know when evaluations should be incorporated into program design, but also where they should not. The United Way that requires an outcome evaluation for every program it funds not only receives a good deal of uninterpretable data, but it also drives people away from their efforts to establish sensible programming. We have to do more outcome evaluations, but we also have to focus these resources where they can make the most impact. In my community, there is a \$20,000 family system services reform grant that requires an outcome evaluation and a \$35,000,000 child welfare program that does not. Evaluators have a responsibility to raise these concerns and help insure that evaluation resources are used wisely, and not with a one-size-fits-all mentality.

Ron Cowell, A Former Policy Maker's View

My comments today come from the perspective of a former legislator and policy maker. The list I made is intended to reflect one person's perspective on some of the major barriers to the effective use of research by policy makers, especially legislators.

First is the absence of a common vision or

Policy makers need the best proximate answer available to the questions they are raising, recognizing that the evidence may be less than conclusive.

purpose. If there is not a recognized common vision or purpose regarding government's role in helping children succeed, it is difficult to agree upon the type of research that can best inform policy, rather than the variety of research that may be used to argue for one or another position.

Second is a political culture that often is not used to examining, let alone critically reviewing, research. Policy makers, and particularly legislators, operate in the context of addressing a variety of issues in a highly charged political atmosphere day after day. It takes a great deal mobilization to gain the attention of policy makers and to get them to sit still long enough to examine research findings, particularly on complex subjects.

Third is the absence of a good communications system about research-based or evidenced-based programs. We need a system that translates these findings into a language and format for policy makers that is credible and understandable.

Fourth is the lack of timeliness of the research. Policy makers who are asking for good information want it now, and they do not find it helpful to hear that they have to wait ten years to have definitive findings. Policy makers need the best proximate answer available to the questions they are raising, recognizing that the evidence may be less than conclusive.

Fifth is that questions that are important to

researchers are not necessarily the same as the ones that are of importance to policy makers. Within the academic community, there needs to be greater emphasis upon conducting policy relevant research, with more interactions between researchers and policy makers.

Sixth is that some of the terms we use do not have an agreed-upon meaning. We need to be more clear and precise about terms such as prevention, early intervention, and family support, or we risk misinterpretations about both agreements and disagreements reached in the policy arena.

Seventh is that we often continue to support programs and activities that have not shown themselves to be effective in producing outcomes for children and families but have a strong constituency. When we support programs with research questioning their effectiveness, it undermines the arguments we are making for the use of research in policy making.

Eighth is that the political system inevitably experiences political transitions, and new Governors or legislatures will want to put their own stamp on the political process. Even programs with demonstrated effectiveness can be abandoned, if they are associated with the former administration.

None of these barriers are insurmountable, but they must be recognized and addressed, if research is to play a greater role in policy making and funding. Polls show that the

Research has to reduce uncertainty for policymakers.

public is willing to invest in children's programs, even if this involves additional taxes, provided the programs will produce results. Researchers have a powerful challenge in convincing legislators and other policy makers that the public's desire for outcome accountability can be satisfied and those investments made.

Christina Groark, A Researcher's View

I would like to stress several points about research and its relationship to policy and practice.

First, research has to reduce uncertainty for policymakers. Researchers cannot overwhelm policy makers or practitioners with data and then equivocate about recommendations and simply call for more time and funding for additional research. When policy makers hear researchers call for additional research without beginning to answer their questions, they quickly become jaded to the research community. Researchers have a responsibility to figure out, with policy makers and practitioners, how to use their findings – to do something in the real world with the information they have gathered and analyzed.

As an example, we know that early childhood programs get results when they are quality programs – when they have trained and supervised staff, when they have good staff-to-child ratios, when they actively involve parents, and when they have good

and broad-based curricula. We can be unequivocal in recommending that these elements be incorporated into early childhood programs, if policy makers want assurances that those programs will produce the gains in child development that we know are possible. We may not have all the answers of how to effectively implement high quality programs in all contexts and situations, but we know enough to reduce uncertainty for policy makers in this area.

Second, policy makers should partner with universities to look at current research being conducted and to explore where it can be more policy relevant. Temple, Penn State University and the University of Pittsburgh have developed a partnership to share our expertise with the state on issues that are relevant in early childhood and school readiness. It is so terrific to have national experts in for seminars like this, but I also believe that Pennsylvania policy makers should make better use of the talent we have in the universities and colleges in the state.

Third, we need to allocate more resources to develop the “learning loop.” We need to promote technical assistance that enhances the use of the information and research we have and are developing. We need to sit together – researchers, policy makers, and practitioners — to put this information to better use in programs and policies.

Finally, the state must get rid of categorical thinking and categorical funding. We have

We should let this debate on whether it should be purely research-based science or a broader evidence-based approach percolate in all of our discussion.

heard about that for a long time and have made some progress, but we need to go to the next level. The outcomes we seek for children – health, safety, school success, responsible sexuality, law-abiding behavior, and civic responsibility – are interrelated and interconnected. In large measure, they stem from a child’s sense of self; a realistic hope for a future; and the safety, stability, and support in his or her environment. Research shows that programs successful in improving that sense of self and the safety and stability and support in the environment improve health, safety, and school success and reduce adolescent parenting and delinquency. Funding, however, remains largely categorical and focused upon a single dimension of well being. We need to figure out how we can get not only our funding, but also our thinking, our policies, and our programs to move beyond categorical boundaries in providing what works.

Clay Yeager, A State Administrator’s Perspective

During the discussion, I was reminded of a conference I attended about twelve years ago where a prominent group of researchers and scholars scolded policy makers and administrators for our failure to promote the importance of continued science-based study that could determine what programs are effective and what are not. Twelve years later, the debate is not about whether or not we accept the importance of science-based study, but rather the manner in which

we define what is good science or evidence and what is marginal. I think we have come a long way.

I believe that we should let this debate on whether it should be purely research-based science or a broader evidence-based (or even the more anecdotal, “it works because we’ve seen it work”) approach percolate in all of our discussion.

I also think that we should recognize that the task ahead is a major one.

Pennsylvania’s annual budget is around \$24 billion. In this legislative session, there is consideration of a Governor’s proposal to set-aside \$22 million specifically for replicating research-based programs. There are efforts to incorporate a stronger outcomes framework into a variety of other initiatives, but there is still a huge share of the budget that is funding programs around which there is uncertainty of effectiveness, either by research-based or evidence-based standards. The task ahead through better research and better allocation choices based upon that research, is to reduce that figure in the years ahead.

CHAPTER 7

Summing Up: Points of Consensus on Using Research in Government Decision-Making

Charles Bruner

I am going to try to synthesize the consensus points that emerged from the various presentations.

First, there was consensus that, where appropriate, randomized clinical trials do provide the most compelling evidence of program impact. We should strive to build this research base but we should also recognize its limitations. In particular, we should recognize that most research-based programs cannot be replicated on a “cookie cutter” manner, but require some adaptation to local conditions and require great attention to implementation that retains high fidelity to the original model.

In addition, while we have a growing number of programs that have been subject to such trials, the current array of proven, “research-based” programs does not cover the range of populations, presenting conditions, or desired impacts and opportunities for action that we would like to produce through programming. In many areas, we still are in the process of developing effective programs and we should not stop these efforts under the name of only supporting proven research-based efforts.

Second, there was consensus that there exists a set of attributes of effective practice that, when integrated into programs, produces more positive impacts. Generally, we should seek to incorporate these attributes into our systems serving vulnerable children and families, including prevention and early intervention programs. Further,

there was consensus that there is an impressive research base regarding these attributes of effective practice, particularly if you look at a broader base of evidence than randomized trials alone. These attributes of effective practice include being holistic, family-focused, and individually tailored, which all require a great deal of worker discretion. Therefore, they cannot be routinized within a specific program curriculum or a fixed intervention protocol. Programs with a strong emphasis on such individualization, where both the intervention and the outcome may be different for different subjects, represent particular challenges to controlled trials. When we seek system or community-wide changes as well as individual ones, the challenge is even greater to constructing an appropriate evaluation strategy

Third, there was real consensus that ours is not a black-and-white world, and we are seeking more than an answer to the question, “Did a program work or not?” We need to know whom it worked for, in what respect, and within what context. We also need to know how much it worked and how significant that is. We have to make quantitative and qualitative judgements on whether the type of impact we are making in the lives of children and families is sufficient to warrant the investment made, compared with other places we might be making the investment. The determination of what constitutes a significant impact extends beyond a determination of statistically significant measured effects and

It is more work to learn the language of evaluation, but in the long run it is a language practitioners need to learn and speak.

requires an assessment of the value of the short- and long-term measured effects and their relationship to program cost.

Fourth, there was consensus that there are a large number of programs that have not shown much in the way of producing a significant impact upon the children and families they serve, however well-meaning they are. Someone has said that we want to avoid reinventing the wheel; we certainly do not want to reinvent the square. We must acknowledge the fact that some programs sound good and make sense, but either because of failed design or inconsistent application and implementation, have not provided evidence that they work. We cannot continue to support such programs and simply hope they will do better. We must challenge them to critically assess themselves and propose and justify programmatic changes that will produce greater opportunities for success.

Fifth, there was consensus that researchers and funders and practitioners all need to become learning partners, guided by data. We have to create tools so that organizations can engage in self-evaluation and practice and learn as they go. These tools must be appropriate to the program, so that we are not recommending hundred thousand-dollar evaluations of thirty thousand-dollar programs. These tools also must be useful to science, based upon attaining the most conclusive evidence of program impact we are able to achieve. Further, we must give attention to fostering innovative prac-

tices that can fill some of the big gaps in our knowledge of what works to help children and families, where we have populations and conditions where we do not really have a very good idea about what works but where we know there is a problem or we know there is a need we have to address. We have to be more concerned about building our knowledge base not only by careful, major research studies, but also through better collection of lessons learned and discoveries made in the thousands of worthy new programmatic efforts being undertaken in the field today that may not be part of a national research effort.

This means doing more work with the research community, drawing upon researchers and evaluators in community colleges and universities. I can imagine some of you pulling your hair out as you think to yourself, “I already have been told I have to collaborate on the community level with child welfare, criminal justice, education and other service systems. Now I am being told I have to collaborate with some researcher from some university. How much more collaboration and learning others’ roles do I need to do?”

It is more work to learn the language of evaluation, but in the long run it is a language practitioners need to learn and speak. Further, the learning of languages needs to be a two-way street. Lisbeth Schorr has quoted Sister Mary Paul as saying that a good programs never tells someone, “We know this is what you need,

A good evaluator should never say to a program, “I know that is what you do, but that is not what we measure.”

but this is not what we do.” We know all too well that that happens too often within our categorical service systems, but effective programs do find ways of meeting what are the essential needs, regardless of their category. This is one of the attributes of effective practice.

I cite that quote because I like to paraphrase it and apply it to the research world. A good evaluator should never say to a program, “I know that is what you do, but that is not what we measure.” When a program describes what it is trying to achieve the research and evaluation community has a responsibility to help construct ways to measure and assess whether it is successful in that effort.

This leads us to my sixth and final point of consensus, around what Chris Groark said. Research has to reduce uncertainty for policy makers and for programs. We must convey research results in a way that really does help improve practice and move us all toward better results for children and families.

Glossary of Terms

Note: The following glossary of terms was provided to symposium participants in preparation for the symposium. Both in the original presentations and in editing them for publication, there was an effort to employ this terminology, where participants made reference to these specific concepts.

Research-based programs. Programs with credible research showing they produce positive, sustained impacts upon the individuals they serve. The criteria for determining which programs have met such a test generally include some or all of the following: (1) rigor of research design, usually requiring an experimental research design with a control group; (2) evidence of measured, statistically significant effects upon individuals served; (3) replication in multiple sites, with equivalent evidence of impact; and (4) evidence the effects are sustained over time. A number of different lists of research-based programs have been developed by different organizations in health, mental health, education, juvenile justice and delinquency prevention, often employing different standards or criteria for inclusion. The Blueprints for Violence Prevention Model Programs employ one of the strictest, or highest, standards for inclusion of programs as a model, research-based violence prevention program.

Attributes of effective practice. Funda-

mental characteristics of service provisions within a program that are considered as essential to producing significant and sustained effects on the individuals they serve. These are generally applied to programs or services that include at least some ongoing counseling, social work, or treatment component. There are various enumerations of these attributes, but they generally include: (1) an approach that establishes relationships of trust with those being served, (2) an approach that works with the individual in the context of family and community, with a strong family focus, (3) a flexible and holistic approach that identifies and addresses fundamental or critical needs, regardless of categorical service boundaries, (4) an approach that recognizes and builds upon strengths and that believes in and holds individuals and families to high expectations, (5) an approach that sets clear goals for participants and tracks progress in meeting those goals, and (6) strong worker skills, training, and support. Many lists of research-based programs include programs that emphasize these attributes as key to their effectiveness. In fact, particularly in the helping professions of counseling individuals or families, it is recognized that the relationship that the worker establishes with the participant is critical to producing results and programmatic interventions will not be successful without the creation and use of this bond to help participants grow and develop. These attributes of effective practice may be more fundamental to program impact than the actual curricula,

training materials, classroom schedules, or other program content that goes along with the counseling or social work activity. Measuring such attributes of effective practice, let alone constructing randomized clinical trials to determine their efficacy, however, represents a significant research and program replication challenge, which may require different methodological tools than traditionally provided in programmatic research, which is much more geared to intervention uniformity.

Evidence-based programs and practices. Programs and practices with a compendium of information, which may include but is not limited to experimental research designs, that together provide compelling evidence of positive program impact. This may be in the form of case studies, documentation of major participant impacts that can realistically be attributed to program activities, and judgments of expert or professional panels of program impact. It also may consider conceptual or theoretical parsimony. In addition to clinical research, there is a large body of evidence related to the attributes of effective practice in the psychiatric, counseling, and social work literature.

Experimental research designs. Research designs that set out testable hypotheses of program impact prior to program initiation and include a **counterfactual**, or comparison to what would have occurred with participants if the program had not been operating. Usually in social science

and medical research, the counterfactual is a comparison group of subjects similar to program participants whose progress is tracked along with the treatment group. To help insure that the treatment and comparison group are similar, individuals are selected for the experiment and then randomly assigned to either a treatment (program) or control (comparison) group. In medical research, the control group receives either a placebo or an alternative treatment, to insure the program and not the participation in the experiment produces the change (in many medical studies, the control group experiences improvement as a result of believing it is being treated, known as a placebo effect). In most education or human service programs, it is not possible to provide a placebo. In some instances, methods other than random assignment are used to establish comparison groups, which generally are referred to as quasi-experimental designs.

Blueprints for Violence Prevention

Model Programs. Programs determined by a National Advisory Board to be research-based in producing a significant prevention or deterrent effect on violence. Model programs must: (1) have a strong research design, e.g. an experimental design with random assignment, low rates of participant attrition, and reliable and valid measurement; (2) evidence of significant prevention or deterrent effects for delinquency, drug use, and/or violence; (3) replication in multiple sites, with evidence of significant prevention or deterrent effects

in those multiple sites, and (4) evidence of effects that are sustained beyond treatment or participation in the program, generally with evidence at least one year post-treatment.

Program integrity. Distinguishing characteristics of a program that define it and can be measured and imparted to others. In addition to specific program features and curricula and training programs, these may include outreach and engagement protocols, staff selection and training and development systems, and record-keeping and referral systems.

Program replication. This involves establishing a program in a new location, maintaining program integrity in the process.

Model drift. Model drift occurs when programs are established in new locations without maintaining program integrity, modifying or eliminating some of the distinguishing program characteristics. Common in the expansion of public programs is for programs to become diluted, serving more participants without commensurate resources, thereby experiencing model drift. When model drift occurs, prior research findings of efficacy or effectiveness cannot be applied to the new program variants.

Program diffusion. Program diffusion occurs when programs are applied to different populations than those they originally were designed to serve. A program that served delinquent youth effectively might

be applied to pre-delinquent youth or to youth with both delinquency and substance abuse. Research showing the program's efficacy or effectiveness with one population, however, cannot be transferred to other populations; additional research is needed to demonstrate whether or not it is effective with those populations.

Fade-out of program impact. Program fade-out refers to the common finding that the impacts of a program on participants are greatest during and immediately after the program, and diminish, or fade-out, over time. If no program effects can be discerned even shortly after program completion, it is unlikely that significant effects will emerge later. Alternatively, some impacts may endure well after program completion, and some program fade-out is to be expected. In fact, single point-in-time interventions are unlikely to be sufficient in themselves to produce life-long impacts, but are likely to require other interventions over time to reinforce and build upon initial gains. This is one reason to examine systemic as well as individual program changes in order to produce positive impacts.

Contextual or environmental effects. Social programs do not operate in a vacuum, but rather within a larger system of family, neighborhood, and community forces that also help shape individuals across social, emotional, health, and educational dimensions. A program that demonstrates an impact upon its participants in one environment or context may result in different

impacts (or no impacts) in another environment or context. For instance, a program that produces reading gains among children, where children are regularly exposed to reading opportunities within their environment, may not produce similar gains among children where children are not so exposed. This is particularly likely with respect to sustained program effects and small program fade-out effects. Again, a pre-school program that produces gains among its participants in pre-literacy skills that operates in a community with poor schools with high student teacher ratios in the elementary grades and limited attention to reading is likely to experience substantial program effect fade-out by third grade, compared with a pre-school program in a community with schools that provide great attention to reading in the elementary grades. In many instances, programs can demonstrate impact (and particularly long-term impact) only if they operate in environments where other supports are in place.

Program efficacy. The ability to produce impacts under ideal circumstances. Research generally seeks to test program impact under controlled conditions, with highly skilled and motivated workers operating programs and participants screened to rule out those who will not be able to complete the program. Research trials under optimal conditions are necessary to determine whether the program can produce results, but these same results cannot necessarily be expected when broadly

applied under real-world conditions.

Program effectiveness. The ability to produce impacts when broadly applied under real-world conditions. Program integrity and program replication without model drift are critical to insuring that programs demonstrating efficacy are able to maintain a similar level of effectiveness as they are exported and expanded.

Statistically significant. The determination, through statistical tests, that the measured program impact could not have occurred by chance. In reference to comparisons between treatment and control (or comparison) groups, statistical significance means that there is an insignificant possibility (e.g. less than a 5% or 1% chance) that measured differences in impact between treatment and control groups could have occurred as a result of chance, e.g. as a result of happening to have assigned participants who would have done better without the program into the program group. The larger the size of the treatment and control group, the smaller the size of a program's impact has to exist for a program impact to be determined to be statistically significant.

Substantively significant. A determination that the actual size of the impact of a program is of sufficient magnitude and importance to warrant attention and to consider be considered efficacious. This requires a considered judgment of whether the measured impacts constitute gains that

are worth the effort undertaken by the program. Programs can have statistically significant impact measures that are of such small size or importance that they would not be considered substantively significant/important.

Pre-post test design. A research design that measures participant condition, performance, or behavior both prior to the program intervention and after completion of the program, to determine if change has occurred during that period. Most experimental designs use a pre-post test design, with the pre-test designed to insure that the treatment and comparison group are, in fact, equivalent. Most programs wishing to develop their own evaluations, however, use pre-post test designs to measure potential program impact, without a control group. The limitations of pre-post test design models are that they usually do not have a good counterfactual to assess what changes would have occurred if the program had not existed. A pre-post test of a program for parents with newborns that measured amount of parental sleep deprivation would find, through a post-test four years after the birth of a child (in the absence of a second birth), that parents show significantly less sleep deprivation among program parents. Such a change, however, would occur without a program. The fact is that time often heals, particularly for people experiencing stress or displaying some risk-related behaviors or conditions.

Proximate and distal effects. Fre-

quently, program goals are to improve general well-being and performance — to improve school performance and completion, to improve health, to improve social behavior. In terms of child outcomes, the public generally wants programs to insure that children grow up safe, healthy, successful in school, and socially and sexually responsible (sometimes measured by the absence of: child abuse, morbidity and mortality, school dropout, delinquent behavior, substance use and abuse, and adolescent parenting). These, however, often represent long-term or distal effects that, particularly for prevention programs, will not be determined for years. Program impacts are likely to be on more near-term, or proximate, measures, such as parental nurturing, provision of primary and preventive health care, age-appropriate cognitive and social behavior, and absence of aggressive or anti-social behavior. These are proximate effects that then must be causally connected to distal effects, if a distal effect is to be posited. There is a growing understanding and knowledge base related to proximate conditions that predict distal effects. In the prevention world, these often are referred to as risk or protective or resiliency factors.

Risk factors, protective factors, and/or resiliency factors. These are factors that have been demonstrated to predict the presence or absence of future problems or social concerns — particularly problems or social concerns related to child development and future likelihood's of school

problems, anti-social and delinquent behavior, adolescent sexual activity and pregnancy, substance use and abuse, and mental illness. The research literature has identified a number of common risk, protective, or resiliency factors (at individual, family, and community levels) that help predict the likelihood of experiencing such distal effects. These factors have been used to develop risk-focused prevention efforts that seek to build resiliency or protective factors or reduce factors in program participants' lives.

Logic model/theory of change. Logic models or theories of change have been used to help programs and systems change efforts conceptually link the proximate effects they seek from their interventions with the distal effects they would like to impact. Logic models and theories of change generally draw from the resiliency and risk and protective factor literature in making these connections.

Cost benefit, cost effectiveness, return on investment, payback period. These are various terms and measurement methodologies used to provide an assessment that a program's impacts produce benefits that can be economically compared with program costs. Program costs usually can be measured in a fairly straightforward manner, but it is more difficult to measure program impacts in terms of dollars and cents. Most of the efforts to develop program cost benefits and assess cost effectiveness have involved examining averted

societal costs as a result of fewer subsequent expenditures on delinquent and criminal activity (and the economic loss of crime itself), reduced welfare dependency, and reduced need for compensatory and rehabilitative services. Particularly for prevention programs, this may require tracking of program participants over long periods of time. Some programs, like the Perry Pre-School program, have quantified the economic benefits and compared them with program costs to calculate impressive overall returns on investment. In addition to measuring ROI's, however, it also is important to examine payback periods or the length of time needed to obtain a positive return on investment. The Perry Pre-School program required more than ten years to payback its initial investment, with its large returns only occurring after participants had reached adulthood. In the business world, both ROI's and payback periods are generally considered together in making decisions

About the Authors

Lisbeth Schorr is Director of the Harvard University Project on Effective Interventions. She is author of *Within Our Reach: Breaking the Cycle of Disadvantage* and *Common Purpose: Strengthening Families and Neighborhoods to Rebuild America*, books which have described effective programs and strategies to improve outcomes for children, families, and communities.

Mark Greenberg is Director of the Prevention Research Center for the Promotion of Human Development at Penn State University. Dr. Greenberg has been a national leader in identifying research-based programs that improve the social, emotional, and cognitive competence of children and youth.

Michael Little is a Senior Research Fellow at the Chapin Hall Center for Children and Bristol University in Great Britain. Much of his work has looked for sustainable connections between good science, policy, and practice in Great Britain. His most recent book (with Kevin Mount), *Prevention and Early Intervention with Children in Need*, draws upon cross-national studies.

Heather Weiss is Founder and Director of the Harvard Family Research Project in the Harvard School of Education. She edits the *Evaluation Exchange*, a publication designed for both researchers and practitioners on evaluation practices that can better assess programs and practices serving children and families.

Cynthia Guy is a Senior Research Association at the Annie E. Casey Foundation, where she manages grants designed to develop and improve research methods and data sources for comprehensive reform efforts. She has managed evaluations of Foundation-funded initiatives in the areas of adolescent pregnancy prevention, employment, and community development.

Charles Bruner is Director of the Child and Family Policy Center, an organization designed to better link “research and policy on issues vital to children and families.” He has written widely on collaborative practices and heads the National Center for Service Integration Clearinghouse.

Tom Gamble is Director of the Civic Institute at Mercyhurst College. Prior to joining the faculty, Dr. Gamble spent fifteen years working in Pennsylvania’s public child welfare and juvenile justice systems. He has co-chairs the Outcomes Workgroup of the Family Services Systems Reform effort.

Ron Cowell is President of the Education Policy and Leadership Center in Harrisburg, Pennsylvania. He served twenty-four years as a member of the Pennsylvania House of Representative and has been active at the national as well as state levels on educational policy issues.

Christina Groark is Co-Director of the University of Pittsburgh Office of Child Development. Dr. Groark has designed and implemented a variety of demonstration

projects, needs assessments, and research studies focusing on young children and their families, including several federal multi-site projects such as the Comprehensive Child Development Program, Healthy Start, and Early Head Start.

Clay Yeager is Director of the Center for Juvenile Justice Training and Research and Executive Director of the Governor's Community Partnership for Safe Children in Pennsylvania. He has devoted his career to the juvenile justice system and oversees Pennsylvania's Communities that Care Initiative, which focuses on training community leaders on risk-focussed prevention efforts to reduce youth violence.

Shawn Carter served a summer internship at the Child and Family Policy Center and completed the initial transcribing for the proceedings.

Deana Brickles, National Center for Service Integration Clearinghouse Coordinator at the Child and Family Policy Center, formatted the publication.

Charles Bruner, Shawn Carter, and Deana Brickles worked with the presenters on reviews and edits of the publication.

Both Centers gratefully acknowledge the support and funding of the Annie E. Casey Foundation for the production of this publication. The opinions and conclusions expressed herein are solely those of the authors and should not be construed as representing the opinions or policy of the Annie E. Casey Foundation.