Challenges and Issues in Designing Applied Research

Robert B. McCall and Christina J. Groark
University of Pittsburgh Office of Child Development

Applied research in child development may be defined as the utilization of behavioral and social science theory and data to describe, explain, and optimize the course of child and adolescent development and to enhance the key settings within which young people develop (e.g., families, schools, after school programs, community social service settings, or health settings). In more common parlance, it deals directly with how we bring "science to life for the benefit of children, youth, and families."

Conceptually, one can imagine a continuum of scientific activities beginning on the left with *"basic" or "foundational" research*, the focus of which is usually the study of cause-and-effect between basic theoretical constructs on the one hand and fundamental and broad-based outcomes on the other.

Moving to the right on the continuum, one can imagine *relevant research*, in which the questions posed are usually grounded in theory and have some relevance to practice. For example, much of the literature of the effects of television violence on children's behavior is "relevant to" but has not prescribed what actions the television industry, policymakers, or parents should take to minimize the potential undesirable consequences of children viewing a steady diet of violent programming.

Moving further to the right is *applied research*, in which the questions posed and answers derived from the studies have rather direct implications for what might be done in practice. The classic intervention demonstration project is a good example, in which an intervention is tried out under relatively controlled conditions to determine if it produces a specific desirable outcome; if so, presumably the intervention might be considered for more widespread implementation in communities conducted by practice professionals whose role in society is to deliver such services (e.g., social workers, teachers, nurses, paraprofessionals, early care and education care givers, etc).

At the far right on the continuum is *practice research*, which consists of studies of phenomena and interventions as they naturally occur in society with the aim of understanding how things work and could be improved. After a new early care and education curriculum, for example, is demonstrated by specially trained teachers perhaps in a lab school to produce positive outcomes, it is desirable to "bring the intervention to scale" in communities where it will be implemented by staff who typically deliver such services in society. Most research on the right side of the continuum is applied research, but much of what this chapter deals with are the issues and challenges of practice research, much of which consists of *program evaluation*.

Main Themes

This chapter has several main themes:

- *Applied and certainly practice research methodology has some unique challenges.* While applied and practiced research methods include elements of basic research methodology, they also face unique challenges and therefore require some unique approaches.
- *The Gold standards of basic research methodology have certain limitations when used in applied and especially practice contexts.* Because gold standard methodology is often viewed as the ideal, applied and practice researchers may confront reviewers of grants

and articles and tenure review committees who view applied and practice research as inferior and methodologically inadequate. While the methodological criterion for basic research is "methodological perfection," the criterion for applied and practice research may be "best obtainable," and for practitioners and policymakers who need to make decisions today the criterion may be "best available."

- *More than one approach is likely necessary in applied and practice research.* By definition, most practice research is conducted in the field, which is an imperfect laboratory, and any single study is likely to have one or more serious methodological limitations. Consequently, several different studies using different methodologies may need to be conducted before results converge on a conclusion that can be persuasively supported.

Much of what is presented below is based on our personal professional experience. One of us (RBM) has spent 40 years conducting basic, relevant, applied, and practice research, and he has seen social and behavioral research evolve from a near total preoccupation with basic research and the experimental method to a discipline that is much broader and more receptive to applied and practice research and the methods appropriate to its pursuit. The other of us (CJG) has had an extensive career in services and policy development with respect to children and families, especially in the development and implementation of new service programs conducted in the community by service professionals. She is well aware of the importance of motivating and organizing policy makers and service professionals to implement a new program smoothly and effectively.

As co-directors of the University of Pittsburgh Office of Child Development, one of the most comprehensive and applied University centers devoted to the welfare of children and families, we have collaborated for two decades with policy makers and community professionals to create and implement new service programs and evaluations in community settings. Most recently, we have collaborated with colleagues in Russia and in Latin America to create intervention programs that comprehensively change the entire operation of an orphanage to increase the warm, sensitive, and responsive caregiver-child interactions by fewer and more stable caregivers in a more family-like, rather than institutional, atmosphere. As a result, the development of infants and young children has been substantially improved in every behavioral and physical domain assessed (St. Petersburg-USA Orphanage Research Team, 2008). Much of what follows rests on this foundation of experience.

A Brief, Personally Interpreted, Methodological History

It helps to understand how social and behavioral science, especially psychology, came to emphasize basic research and the traditional gold standard-methodology (see below; Groark & McCall, 2005).

In 1945, Vannevor Bush argued for a "social contract" between science and society based on the assertion that all scientific knowledge is potentially useful, at least someday, implemented by someone, for some purpose. Thus, scientists should be guided by their curiosity to study phenomena of conceptual and theoretical interest, with the assumption that it may be useful someday in some way.

In the next few decades, psychology and other behavioral disciplines were trying to establish a scientific identity; they wanted to be "sciences" like chemistry and physics. To counter these accusations that they were not a science and their empiricism was "soft" or worse and to establish scientific credibility, they adopted research methodologies that were analogous to those in the "hard disciplines," such as theory-driven research, random assignment of subjects to conditions, uniform treatment administration, automated or blind assessments, and statistical analysis of data (see below). These gold standards were used to pursue general cause-effect laws of behavior that were presumed to explain most behavior in most contexts, so studies of gender, individual differences, context, or procedures were presumed to be unnecessary.

At the same time, government was willing to fund basic behavioral research, so the confluence of these historical, social, and economic themes was that basic research and the gold standard methodologies became the coin of the behavioral academic realm. Indeed, this value system became the criteria for judging grants, publications, and academic tenure.

However, after the principles of basic learning were articulated (e.g. reinforcement, generalization, discrimination, extinction, etc.), general cause and effect principles of behavior became more elusive; cause-effect phenomena often seemed to be qualified by "nuisance factors" such as gender, individual differences, minor procedural variations, wording of instructions, particular stimuli and contexts, one vs. another outcome measures – indeed, every aspect of the research process. Moreover, Bush's social contract began to wear thin in the halls of Congress, which demanded more relevancy, application, and practical value for their behavioral research dollar.

The net result of these themes has been that behavior scientists now study a great deal more complicated phenomena and more applied and practice topics. But while these issues present new challenges and often require different methodologies, the value for – often insistence upon – traditional gold standard methods persists and is often invoked – to an inappropriate extent (in the authors' opinions) – as criteria for judging applied and especially practice research.

Benefits and Limitations of Gold Standard Methodology for Applied and Practice Research

The gold standard methods were developed to primarily demonstrate, as unequivocally as possible, cause-and-effect relations between some independent variable *X* and some dependent variable *Y*. This criterion is called *internal validity*. Applied and practiced research are certainly concerned with internal validity; for example, *can* curriculum X be demonstrated to produce better school readiness skill Y? But they often have an additional concern called *external validity*: *Does* curriculum X produce school readiness skill Y in naturalistic conditions (e.g., when curriculum X is implemented by teachers for students in schools that are typical of a given community)? Logically, it is difficult to obtain external validity without internal validity, but it is possible to have internal validity without a great deal of external validity. Experimental interventions, for example, may be demonstrated to be effective under random assignment and carefully controlled conditions but are difficult to implement successfully in community contexts. Also, the better a method is at demonstrating internal validity, the poorer it tends to be at demonstrating external validity, and the reverse.

What follows is a brief discussion of the merits and limitations of several gold standards when implemented in applied and practice research. Please note that what follows is not an argument against the gold standards; it is a discussion of their assets and especially their limitations, which are often ignored, when used in applied and practice research contexts.

Research methods are tools, such as a hammer and saw, each of which is better suited for one than another purpose (for more complete discussion see McCall & Green, 2004).

Theory-Driven Research

Standard: *Research should be guided by and contribute to theory.*

Theory usually refers to a network of interrelated causal principles (e.g., learning theory, psychoanalytical theory), but it can also refer to a single causal principle.

Benefits. Theory describes causes and effects, it usually applies to many different circumstances, it predicts to unstudied new circumstances, it explains and makes phenomena understandable, and to some (Weiss, 1995) there is "nothing so practical as a good theory." For example, a single theoretical principle is that adolescents who perceive that they have a realistic chance at a successful and fulfilling future are less likely to engage in adolescent problem behavior (e.g., excessive alcohol and drug use, risky sexual behavior, poor school performance, antisocial and criminal behavior). This principle predicts that if low income adolescents are guaranteed college tuition, they will do better in high school and engage in less problem behavior. The principle makes understandable why guaranteeing tuition has this effect, and there are many other ways to providing a realistic future that fall under the principle. Theory is so standard that it is difficult to get a grant approved or an article accepted without providing a theoretical context for the question to be studied.

Limitations. Not all research that is of applied and practical value has a theoretical context. Applied research often focuses on the detection and description of certain problems or circumstances that have important behavioral and financial consequences for individuals in society. Topics may be studied, not for their theoretical relevance, but because of their personal and social costs. It is useful to know whether school dropout rates are increasing or decreasing, which groups of people have high rates, what school circumstances and other parameters are associated with high rates. This information can help target attention and resources to deal with the problem, and perhaps eventually to contribute to theoretical explanations.

Also, some social programs are implemented for political, philosophical, or religious reasons, such as abstinency programs to prevent unwanted pregnancy among teenagers. Nevertheless, it is important to know whether such programs indeed lower teenage pregnancy rates, regardless of whether there are theories about how or why.

Agreed, it is always helpful to know the causal mechanisms that make an intervention successful, because such information tells you what can and cannot be changed to fit the program to local circumstances. *Program evaluation* is primarily oriented at determining whether an intervention is faithfully implemented and produces the intended outcomes, whereas *program evaluation research* is aimed additionally at discovering crucial characteristics and possible mechanisms that produce those benefits.

Note that the absence of theory does not justify "random research." There needs to be a rationale for implementing and evaluating an intervention and for why it should be effective, but that rationale may be economic (large amounts of money are already invested in the program) or human and societal costs involved.

Random Assignment

Standard: *Participants should be randomly assignment to experimental/treatment versus control/comparison groups.*

Benefits. Random assignment is perhaps the shiniest of the gold standards, because randomizing participants to treatment groups randomizes subject characteristics and self-selection factors across experimental conditions, thereby contributing strongly to the internal validity of cause-and-effect. Further, a double blind design with a placebo condition in which neither the participants nor the data collectors are aware of which participant is in which group provides a further check on participant as well as observer biases. Drug trials are the classic example, but note that participant commitment to the intervention is not required.

Limitations. Sometimes participants cannot be randomly assigned to certain conditions. Children cannot be randomly assigned to divorced versus widowed single mothers, for example, or to father absent versus present families. Further, when behavioral interventions are involved, it is difficult to prevent participants from knowing which treatment group they are in, in which case double blind (and even single blind) studies are impossible. Finally, in contrast to most drug trials, participant engagement and commitment to the intervention may be crucial to its success, and people may drop out of interventions because they don't like it and those in the comparison group may obtain the same or a similar intervention on their own (see below the section on intention to treat analyses). Or participants may remain in their assigned group, but be disappointed they were not in the other ("reactive disappointment"), which can dilute the effectiveness of the treatment.

These kinds of problems are common in practice research. For example, low-income families with a newborn child were randomly assigned to the Comprehensive Child Development Program treatment versus the control groups, but ethnographers discovered that control subjects benefited from the information and services provided to their treated friend or relative (McAllister, 1993). Further, schools were randomly assigned to Comer's School Development Program (Comer, 1988; Comer, Haynes, Joyner, & Ben-Avie, 1996; Joyner, Comer, and Ben-Avie, 2004) and to a no-treatment comparison group. But in one evaluation (Millsap, Chase, Obeidallah, Perez-Smith, Brigham, & Johnston, 2000), there was little difference between the two groups, primarily because some schools in the treatment group did not implement the treatment very vigorously and some schools in the comparison group implemented elements of the program on their own. However, positive correlations were found within both groups between the extent to which a school implemented the principles and the beneficial outcomes. Do we conclude that the program did not work because there was no difference between the randomly assigned group means, or that it did work because schools that chose to vigorously implement it produced better results (see intent to treat section)?

It is also sometimes said that random assignment produces the maximum treatment effect size, but this is not necessarily the case. In the Comer example above, the randomly assigned treatment effect was nearly zero, but the correlational results suggest if schools were allowed to choose which approach they wanted, the treatment effect for the Comer program would have been much larger.

Nevertheless, the primary advantage of random assignment is that it minimizes the effect of subject characteristics, especially those correlated with self-selection into treatment groups. So in the Comer study above, it is fair to ask whether *any* program that schools have faith in and commitment to implement would have the same benefits. Other programs may be much cheaper and easier to implement and produce some or even more benefits.

Strategies. One approach to solving this problem is to use a design that includes both random assignment and self-selection. For example, suppose one wanted to know whether mediated custody arrangements produce better child adjustment to divorce than court ordered arrangements. In this case it is obvious that the divorcing parents need to be committed to the decision process, and some parents are likely to do better under mediation and others under court orders. So have one condition that randomly assigns couples to the two arrangements whereas in another condition parents are allowed to choose, then measure in both groups as many characteristics of couples that are likely to predict (i.e. correlate with) a couple's self-selected choice and validate these predictors within the self-selection group. This design can assess directly the effect of self-selection and participant commitment by comparing that condition with random assignment. The random assignment provides the opportunity for internal validity support, and the self-selection condition may mimic what will be offered to couples in the future (i.e., external validity). This example also illustrates that when services are offered on a routine basis within the community, participants are rarely if ever randomly assigned (except, perhaps, for certain court ordered interventions). A random assignment study may provide internal validity but very limited external validity, since one may be uncertain about the generality of results from a random assignment study to community conditions in which participants choose whether and which service to partake.

Uniform Treatment

Standard: *Uniform treatment implementation in which the experimenter controls the treatment vs. no treatment experience of participants and each participant in the treatment group is administered the same treatment.*

Benefits. Uniform treatment permits the researcher to know exactly what treatment vs. control conditions produced the observed results, it reduces error variance otherwise associated with variability in treatment implementation conditions, and it permits replication of the treatment or intervention.

Limitations. This strategy assumes that one intervention fits all. But some treatments are deliberately not uniform, such as family support services that are tailored to the specific needs of each individual family. Further, it is sometimes difficult to prevent service providers from tailoring a treatment to fit participant characteristics and circumstances, which is a hallmark of good treatment practice. Also, participants in interventions and especially control groups cannot practically or ethically be prevented from seeking a variety of treatments on their own, including control participants who obtain elements of the experimental treatment or other services that may compete, supplement, or interfere with the experimental vs. control effects.

Strategies. An approach to dealing with these complexities is to measure the aspects of the intended as well as unintended treatment activities and services in *both* treatment and comparison groups. Assuming such measurements are at the individual level, hierarchical linear modeling (Bryk & Raudenbush, 1992) can be used to test the randomly assigned treatment vs. comparison group difference and the treatment implementation characteristics reflected in the participant level variables imbedded within those groups.

Quantitative Measurement and Statistical Analysis

Standard: *Measure independent and dependent variables quantitatively and analyze the data statistically.*

Benefits. This strategy allows the researcher to distinguish and quantify characteristics of the treatment and treatment effects, and to make judgments that distinguish estimated treatment effects from random variation. It also allows the researcher to communicate in quantitative terms the extent of reliability and validity in measures, and the treatment effects.

Limitations. Quantitative measurement assumes we know what and how to measure before starting the study. When studying new phenomena, this may not be the case. Of course, we have hypotheses and guesses, but sometimes those are incorrect (e.g., searching for lasting IQ gains as a result of early care and education programs rather than school achievement and antisocial behavior; McCall, Larsen, & Ingram, 2003). Datta (1994) suggested that rather than dropping a fishing line with a worm into an unknown lake trying to catch a new species of fish, one needs to first send a scuba diver to observe where fish congregate and what they eat. Further, some phenomena may be difficult to quantify, and qualitative, perhaps ethnographic, descriptions might be more informative, such as when studying how gangs are started, maintain order and cohesiveness, promote member loyalty, and maintain lines of authority; or the process of community service systems change. Although the debate has quieted down, the quantitative-qualitative discussions of past decades may be instructive (Reichardt & Rallis, 1994).

Further, many statistical techniques commonly used by psychologists may be less appropriate for research in which practitioners and policy makers are an important audience. For example, averages mask extreme cases, but extreme cases are often the focus of services and justify financial appropriations. Also, effect sizes expressed in terms of percent variance are less useful to policy makers than statistics that communicate the rate of service needs reduced by the treatment and cost savings (see Scott, Mason, & Chapman, 1999).

Strategies. Ethnography might be more descriptive, especially when studying new and complex phenomena, both for the purpose of generating hypotheses as well as obtaining initial answers. Originally developed to study single cultures (n=1), ethnography can also be used to study subcultures within a society, and modern ethnographic and conventional statistical procedures can be used jointly in single studies. Also, qualitative procedures can often explain results. For example, a program aimed at providing services for teenage mothers who had drug and alcohol problems found that the average age of participants was 26 years, not the intended 16-18. Simply asking the service staff why this was the case produced the hypothesis that teenagers do not believe they have a problem because so many of their friends also use drugs and alcohol, whereas 26-year-old unmarried mothers have come to realize they have a substance abuse problem and are not being good parents. The service organization then used the 26-year-olds to counsel the teenagers to participate in the service program.

Conclusion

The gold standard methodologies are ideally suited for basic research and demonstration projects in which high internal validity of cause-and-effect is desired. They have many advantages that are just as necessary in applied, but they also have limitations that become more prominent when these methods are used in applied and practice contexts. Consequently, other methodologies are also needed as complements or sometimes necessary substitutes, which attempt to handle the challenges of applied research but in turn have their own limitations. Conclusions of internal and external validity can be inferred from converging evidence from a set of studies, each of which has complementary assets and limitations.

Issues in Working in the Community

Academic-Community Collaborations

Applied and especially practice research often require that the study be conducted in community contexts (e.g., human service agencies operating in the community, schools, hospitals, early care and education facilities). These organizations and their directors and staff have different values, purposes, procedures, regulatory constraints, constituencies, and performance criteria than do researchers (Groark & McCall, 1993, 1996). Further, "community professionals" (e.g., agency directors, teachers, physicians) and researchers may possess some degree of distrust, lack of respect, and diminished value for the other group and their activities. While the researcher may control the funding for the research project, the community professionals control everything that happens in their organizations and institutions; researchers must recognize that many practice studies are conducted in the community's "stadium," need to be conducted with them as major players, and operate by their rules.

Consequently, special efforts – and attitudes – may be needed to create and maintain an effective collaborative group of researchers and community professionals to conduct a mutually beneficial project. Collaborations may be difficult for academics and some community professionals as well, because they inevitably involve diminished control by each individual member of the collaboration as well as require compromises that may rub coarsely against traditional standards and values. Researchers considering a collaboration with community professionals may want to consult suggestions on how to create and operate successful research-community collaborations (e.g., Groark & McCall, 1993, 1996, 2005, 2008).

Project creation. Usually, field projects need to be created collaboratively between researchers and community professionals, although there are some exceptions.

Evidence-based programs. Sometimes the state or local funders (e.g., Department of Welfare; Department of Children and Youth Services; local foundations) identify a service program that has been demonstrated to be successful, usually in applied research demonstrations but sometimes more broadly (e.g. Nurse-Family Partnership; Olds & Kitzman, 1993; Olds, Henderson, Kitzman, Eckenrode, Cote, & Tatelbaum, 1999), and want it implemented locally. Indeed, some policy makers and funders subscribe to a narrow definition of "evidence based programming," which largely consists of attempting to replicate locally service programs demonstrated elsewhere to be successful (McCall, Groark, & Nelkin, 2004; Groark & McCall, 2005, 2008). But such a strategy makes several implicit assumptions (McCall et al, 2004; Groark & McCall, 2005; McCall, in press), namely that sufficient evidence with reasonable internal and external validity information is available, the program itself is packaged so others can implement it, others in fact implement it faithfully, and it matches local needs and resources. While policy makers and funders can encourage with financial resources the adoption of a specific service program or intervention, there is no guarantee that the community professionals have the same motivation, commitment, and skills as the creators of the program; that the original program is not modified to fit local participants, circumstances, and budgets and such modifications may harm its effectiveness; or that the program is faithfully or effectively implemented – indeed, implementation of programs is exceedingly crucial to their success but a process that is relatively unstudied and therefore without evidence-based procedures (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005; McCall, in press).

Collaboratively creating programs based on evidence. Even if an evidence-based program is in hand, collaborative planning with the community will be necessary (Groark & McCall, 2008; McCall, in press) especially if a researcher (as opposed to a policy maker who has influence over community organizations) has an intervention in mind or no well-articulated and demonstrated service program or intervention is available. It is unlikely that a researcher can successfully "drop" a pre-planned intervention or program into a community organization without going through some process designed to engage the community professionals, convince them of the merit for them of the proposed project, and garner their commitment and cooperation to implement it and to cooperate with the evaluation, which is often costly in time and in convenience to them. Instead, researchers and community professionals should collaboratively identify needs and create an intervention together, perhaps using a logic model approach (Armstrong & Barsion, 2006; Axford, Berry, & Little, 2006; W. K. Kellogg Foundation, 1998, 2000), which has been packaged in different ways (e.g., Pathways Mapping) (Schorr, 2003) and Getting to Outcomes (Wandersman, Imm, Chinman, & Kaftarian, 1999, 2000; http://www.rand.org/pubs/technical_reports/TR370/; Fisher, Imm, Chinman, & Wandersman, 2006). These different versions of a logic model simply provide a structure consisting of a set of questions that the collaborators need to answer on the way to developing an appropriate program with measures of implementation and outcome that is consistent with research evidence and best practices. Sometimes a knowledgeable but independent facilitator is helpful in conducting the process, which may take one or two sessions or repeated sessions over several months. While there is no research on the outcome effectiveness of logic models, the intended result is the creation of a program that meets local needs, can be implemented with local personnel and financial resources, has the commitment and enthusiasm of all of the major stakeholders, embodies evidence based principles, and includes a monitoring and evaluation plan (which may be a researcher's primary interest and responsibility).

Other Issues

Several other issues arise in field research.

The rush to outcome. Often funders want to know that an intervention "works," and they want to know it soon, which can mean after the first cohort of participants completes the intervention. In other cases, the primary outcome is years away (e.g., school success, graduation, and leading a financially self-supporting life) when the intervention consists of an early care and education program, in which case intermediate goals and outcomes will need to suffice. The expectation of, or demand for, outcomes in the first cohort may be unrealistic, because new and complex programs often require the staff to experience two or three cohorts of participants before the program is implemented smoothly and effectively Fixsen et al., 2005). Often it is helpful to have the funder in the planning and implementation group of stakeholders, and to have this collaborative establish not only a timeline for the implementation of the service or intervention but also a timeline of expected outcomes that represents a compromise between what can be reasonably accomplished and the patience and resources of the funder. Further, the first "outcome" is that the program is implemented faithfully and effectively; only then is it reasonable to expect beneficial outcomes for participants (Groark & McCall, 2008).

Independent vs. Participatory Evaluation

Historically, researchers often evaluated an intervention or service which they created and operated, but when field tests of potential public services became for prevalent, policy makers and funders insisted that the evaluator be independent of the service provider. This often fueled attitudes of skepticism and mild antagonism between evaluators and service professionals, who perceived the evaluators to be testing and grading their performance. More recently, various forms of participatory evaluation have been emphasized (Fetterman, 1993; Fetterman, Kaftarian, & Wandersman, 1996) in which evaluators and program professionals collaborate to plan, implement, and execute a program evaluation.

Both independent and collaborative evaluations have some assets and liabilities. Funders often prefer "independent evaluations," because they provide the evaluation with a measure of credibility over service professionals who are assumed to be highly committed to the success of the intervention. But if the evaluation is conducted too independently, service professionals feel that the evaluation is being "done to them," and the evaluators, who have limited relationships with the service professionals or the participants, may not be able to collect the kind of data on program implementation and outcome that is desired.

Conversely, participatory evaluation can have the benefits of an involved professional staff who can use their relationships with participants to obtain more personal and in depth information on outcome and collect such information at less cost. However, motivating service providers to collect data can be challenging, they often lack the commitment to accuracy and comprehensiveness that researchers and their assistants have, and questions may linger about the validity and credibility of the results if the service providers have too much responsibility in collecting data. Also, sometimes a major stakeholder who is part of a collaborative team is actually responsible for a crucial component of the intervention or data collection (e.g., supervisory staff), and it may be their domain that is not functioning adequately. Such situations need to be handled as soon as possible, forthrightly, but discreetly, perhaps in private meetings between evaluators and this individual.

Of course, there are compromises between these two extremes. Researchers must be clever in training service personnel and designing data collection to make it useful to the service providers, monitoring of data collection needs to occur to insure its accuracy and comprehensiveness, participants can mail to researchers questionnaires given to them by service personnel which they answer privately, etc.

Applied Research Designs and Strategies

In applied and practice studies, the researcher has much less control than in basic research, so comparison groups are often difficult to obtain, participants do not necessarily stay in their assigned groups, budgets and personnel change mid-project, participants come and go at different ages, and so forth. Research designed for such projects may need to be quite different than for basic research and require considerable creativity to identify and obtain appropriate comparisons. A variety of old and recent textbooks and handbooks exist describing different research designs and strategies for these circumstances under the rubrics of "quasi-experimental design," "applied research methodology," and "program evaluation" (e.g., Cook & Campbell, 1979, Rossi, Lipsey, & Freeman, 2004). Newcomers to applied and practice research should become familiar with these strategies so they have an arsenal of tactics that can be creatively selected, combined, or modified to fit the circumstances of any particular applied or practiced study. A few of the more commonly used strategies are outlined very briefly below as an introduction, but advanced specialized references should be consulted before implementing these strategies (see list).

No Comparison Group

Practice researchers will often find themselves without an obvious comparison group. No-treatment controls are difficult and expensive to obtain, it is unethical to deny treatment that may be available to needy participants, there is limited benefit to participation control individuals, and people who are denied treatment often seek it out on their own.

Post-test only. Sometimes program evaluators are called in after a program has been in operation for several years and the only data available are essentially "post-test" or "outcome" data. In this case, nearly the only strategy is to compare treated participants with "norms" for the outcome assessment based on a standardization sample or a very large study available in the literature. For example, children reared from birth in orphanages are sometimes adopted into advantaged homes and parents are asked to respond to questionnaires regarding their children's adjustment. The Child Behavior Problem Checklist (Achenbach & Edelbrock, 1983) is often used, mean scores, and the percentage of children who fall into the borderline and clinical areas on the scales are often compared with the CBCL standardization sample of parent-reared USA children.

Single group, pre- and post-test. Sometimes researchers are able to obtain assessments before participants receive an intervention and again at some point during, immediately after the intervention is completed, or at a follow-up assessment some time later. If age-related norms exist for the assessment instrument, one can determine if participants moved up in percentile ranking or some other standardized score (T score).

However, if norms do not exist, the researcher is faced with the task of demonstrating the changes from pre- to post-test occurred because of the intervention, not because the children increased in age or some other factor that changes with the passage of time. A strategy to deal with this situation (Bryk & Weisberg, 1976; McCall, Ryan, & Green, 1999) can be implemented if children enter the intervention at different ages, even if they remain in the intervention for different lengths of time. A "residualized change score" can be calculated by regressing pre-test scores on age, predicting for each individual subject their outcome score based on this regression using age at post-test as a parameter, and then using the difference between actual and predicted post-test score as the dependent variable reflecting treatment effect.

This approach can be very useful when participants are expected to change on the outcome measure simply as a function of time (e.g., low-income children are known to decline in mental test performance over the preschool years relative to the population). For example, one early childhood family support intervention for low-income children 3-6 years of age assessed general developmental status and found that mean performance actually declined between pre- and post-test. But when children's pre-tests were regressed against age, a negative relation was observed, indicating that such children would decline in performance even in their own homes with no intervention. When residualized change scores were calculated, however, children in the intervention actually declined less than would have been expected without the program, indicating that the program was "effective" in preventing decline (McCall et al., 1999).

Random Assignment with Waiting List Comparison Group

If an intervention has been demonstrated to be successful elsewhere, it may be unethical to deny treatment to individuals in a similar or replicated intervention. One strategy is to use a delayed intervention in which a group of worthy participants is selected, but randomly assigned to receive the intervention immediately or some time later. Both groups are assessed at the beginning, at the end of the first group's intervention, and after the second group's intervention. The delayed group constitutes the no-treatment comparison for the immediate group, and some assessment of longer-term benefits is possible.

Sometimes financial and personnel resources can only treat a limited number of participants, so participants are randomly assigned to treatment now vs. a waiting list (i.e., treatment later) or perhaps no treatment if resources are very limited. Service professionals often want to treat the most needy, but if there are more needy than can be treated, they may welcome random assignment as a "neutral" strategy for deciding who is treated now and who is treated later or not at all.

Two-Intervention Comparison Design

One strategy that avoids not treating some potential participants is to simultaneously implement two contrasting interventions but to conduct pre- and post-treatment assessments of both treatment outcomes on both groups. The question is which intervention provides better results compared to the other, rather than compared to no intervention. The expectation is that Treatment A will affect Outcome A but not Outcome B, whereas the reverse will be the case for participants in Treatment B. For example, two different early care and education curricula, one emphasizing emergent literacy and the other emergent numeracy, might be implemented in different centers, but children are measured both pre- and post-intervention on both literacy and numeracy. Children in the literacy intervention should be expected to improve on literacy to a greater extent than children in the numeracy intervention, but the reverse would be expected for the numeracy children. This strategy only works to the extent the treatments do not generalize to the other outcome.

No-Shows and Dropouts in Randomized Studies

Applied and practice projects, especially those using low-income and high-risk participants, often experience substantial no-show and dropout rates. Low-income families move often, participants do not attend all parts of the intervention, some tire of it, some perceive no benefit and drop out, and some regard it as more hassle than potential benefit. It is not unusual to have only a small percentage (less than 50%-70%) of participants randomly assigned to treatment and comparison groups actually be exposed to the entire intervention or service program.

Promoting participation. A major responsibility of intervention and service programs is to motivate people (which may be both teachers, caregivers, nurses who may implement the intervention on the one hand and client participants on the other) to participate in the intervention and in the control conditions. Not every potential participant wants to participate, especially in random assignment studies, even though researchers and service providers believe that they need it and should be motivated to participate. Incentives may need to be offered to participants in both groups, preliminary sessions may help build enthusiasm for participation, and these early sessions may be conducted before random assignment to weed out those who would quickly drop out.

Intent to treat analysis (ITT). Nevertheless, it is likely that all of those participants assigned to each group will not complete all of the intervention or comparison conditions but may nevertheless have both pre- and post-test assessments. It is often recommended, indeed insisted upon, that an Intent to Treat (ITT) statistical analysis is conducted in which all individuals who were assigned to a group (were intended to be treated) are included, regardless of whether they actually received all or any of the treatment. Such analyses produce estimates based on participants assigned to, not who received, the intervention or services.

Sometimes, ITT is indeed the appropriate analysis for the question at hand. For example, suppose a flu vaccine is offered free to a designated group (i.e., those over 60 years of age) in one community and not another, and the incidence of flu in that group, whether or not they actually received the immunization, is compared between the two communities. This comparison appropriately addresses the question of whether the free immunization reduced flu in the community. Notice the question is not whether the flu immunization is effective at preventing flu. ITT works best if the non-compliance rate is small relative to the total sample, outcome data are available on nearly all who were assigned to both groups, and the intervention benefits do not require much participant commitment.

However, the greater the dropout rate, the less sensitive to detecting treatment effects the ITT strategy becomes, because a progressively larger percentage of those assigned to the treatment actually did not receive all or any of it and cannot be expected to benefit from it. The most obvious response is to take only those participants who actually received all or most of the intervention and compare them to the control group, but critics argue that these are self-selected and more highly motivated people and are not typical or representative of the target population or they might have improved with any program or even no program. Opponents argue that it makes no sense to include those who were not treated. One can calculate with estimates: The ITT result represents a lower estimate of intervention effectiveness whereas the result for those who completed the program represents an upper estimate.

Alternatives to ITT. One strategy is to measure a variety of variables hypothesized to predict those individuals who stay in the treatment condition vs. those who drop out, validate that these measures do indeed predict complete participation, select participants in the control group who match those who stayed in the intervention on the basis of these variables, and analyze the difference between completed intervention participants and their matched subset of control subjects (a type of "propensity score" analysis; Foster, 2003; Imbens, 2000; Rosenbaum & Rubin, 1983). Another simpler alternative is to use the Treatment on the Treated (TOT) approach (Bloom, 1984; Ludwig & Phillips, 2007) strategy in which the TOT impact is equal to the difference in the average outcome of interest for children *assigned* to the treatment vs. control group (the ITT impact) divided by the difference in program *enrollments rates* between the treatment and control group. This strategy is roughly equivalent to the "Complier Average Causal Effect" (CACE) analysis, which works best if one has both predictors of compliance versus dropout and outcome covariates (Frangakis & Rubin, 1999; Jo, 2002; Muthén & Muthén, 1998-2001; Yau & Little, 2001).

Statistical Strategies

Not surprisingly, the complexity in design and assessment strategies that often accompanies applied and practiced research must be analyzed with statistical techniques that match that complexity. Many modern statistical techniques have been designed to deal with these circumstances, including structural equation modeling, hierarchical linear modeling, latent growth-curve analysis, instrumental variables, propensity analysis, and so on. A list of references accompanies this chapter that includes articles that provide readable introductions to these and other statistical techniques plus some that describe how to implement and interpret such analyses.

Obtaining Tenure for Applied and Practice Research

Given the assumption that behavioral science, especially psychology, favors gold-standard methodologies even when conducting applied and practice research, it is reasonable to ask if young scholars can obtain tenure, especially in a traditionally oriented psychology department, if they devote themselves to studying applied and practical problems.

First, this depends on the tenured members of the department, the Dean of the school, and the Dean's school-wide tenure-review committee, all of which must approve the tenure decision. If these individuals are highly committed to basic research and have a low value for and negative attitude toward applied scholarly activities, the going is likely to be rough. Some candidates in this situation make tenure anyway if they generate enough basic research, literature reviews, grant support, grants from the "right" agencies, and publications in the "right" journals to merit tenure in spite of their applied activities. Instead of trying to conduct basic and applied research simultaneously, do it sequentially: Conduct basic research and do what is necessary to get tenure, and then conduct applied and practical research.

In any case, young scholars should try to pick a department that values the kind of scholarly activities they want to pursue. Not all psychology departments are very traditional, departments in professional schools (e.g., social work, education, public health), for example, may be much more receptive, understanding, and rewarding of applied and practice research than traditional psychology departments, although some professional schools believe the road to academic respect and increased prestige is through basic research. Some (McCall, 1996) have argued that the goal of all scholarship is to contribute to the benefit of humanity and other living things, and the distinction between basic, applied, and practice research is artificial in view of this common goal, but rarely do such arguments carry the day.

Some young scholars don't care. They are committed to contributing as directly as possible to the welfare of children, youth, and families, so they pursue their interests and career outside of traditional academics, for example, in a private research organization that specializes in applied research or in a public or private service organization large enough to conduct research and evaluation on services. Such organizations are often pleased to have people with such training, but at least some of these positions bring pressure on young scholars to obtain sufficient research funds to pay their own salary (but this pressure can also exist in universities).

Conclusion

Society needs researchers who study applied and practice programs and topics, Congress has called for such research for many decades, and NIH now dedicates substantial funding to "translational" research. Historically, however, applied research has been denigrated, partly because it is often less precise than basic research, conclusions of a single study are more ambiguous, and frankly some studies are methodologically sloppy and loose in drawing conclusions based on inadequate evidence. Applied and practice research deserves to be denigrated if it is done poorly, but there is a difference between "crude" and "sloppy" research. Society needs to base its policy and practice on the "best obtainable" evidence, even if not perfect, and sometimes even the "best available" evidence at the time the practitioner and policy maker must make their decisions. Because applied and practice research is more difficult, behavioral and social science disciplines need to "care enough to send their very best researchers to the task." While the methodological challenges and profession frustrations may be great, so is the satisfaction and fulfillment of having one's professional activities more clearly contribute to the welfare of children, youth, and families.

References

Achenbach, T. M., & Edelbrock, C. (1983). *Manual for the child behavior checklist and revised child behavior profile*. Burlington, VT: Queen City Printers.

Armstrong, E. G., & Barsion, S. J. (2006). Using an outcomes-logic-model approach to evaluate a faculty development program for medical educators. *Journal of the Association of American Medical Colleges, 51,* 483-488.

Axford, N., Berry, V., & Little, M. (2006). Enhancing service evaluability: Lessons learned from a programme for disaffected young people. *Children and Society, 14*, 287-298.

Bloom, H. S. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review, 8*(2), 225-246.

Bryk, A., & Raudenbush, S. W. (1992). *Hierarchical linear models for social and behavioral research: Applications and data analysis methods.* Newbury Park, CA: Sage.

Bryk, A. S., & Weisberg, H. I. (1976). Value added analysis: A dynamic approach to the estimation of treatment effects. *Journal of Educational Statistics, 1*(2), 127-155.

Bush, V. (1945). *Science: The endless frontier.* Washington, DC: U.S. Government Printing Office.

Comer, J. P. (1988). Educating poor minority children. *Scientific American, 259*, 42-48.

Comer, J. P., Haynes, N. M., Joyner, E. T., & Ben-Avie, M. (1996). *Rallying the whole village: The Comer process for reforming education.* New York: Teachers College Press.

Cook, T., & Campbell, D. T. (1979). *Quasi-experimental design and analysis issues for field settings*. Chicago, IL: Rand McNally.

Datta, L. E. (1994). Paradigm wars: A basis for peaceful coexistence and beyond. In C. S. Reichardt & S. F. Rallis (Eds.), The qualitative-quantitative debate: New perspectives. *New Directions for Program Evaluation Monograph, 61*, 53-70. San Francisco, CA: Jossey-Bass.

Fetterman, D. M. (1993). Empowerment evaluation. *Evaluation Practice, 15*, 1-15.

Fetterman, D. M., Kaftarian, S. J., & Wandersman, A. (1996). *Empowerment evaluation: Knowledge and tools for self-assessment and accountability.* Thousand Oaks, CA: Sage.

Fisher, D., Imm, P., Chinman, M., & Wandersman, A. (2006). *Getting to outcomes with Developmental Assets: Ten steps to measuring success in youth programs and communities*. Minneapolis, MN: Search Institute, www.search-institute.org.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). Implementation research: A synthesis of the literature. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

Foster, E. M. (2003). Is more treatment better than less? An application of propensity score analysis. *Medical Care, 41*(10), 1183-1192.

Frangakis, C. E., & Rubin, D. B. (1999). Addressing compilations of intention-to-treat analysis in the combined presence of all-or-none treatment-compliance and subsequent missing outcomes. *Biometrika*, 86, 365-379.

Groark, C. J., & McCall, R. B. (1993, Spring). Building mutually beneficial collaborations between researchers and community service professionals. *Newsletter of the Society for Research in Child Development,* 6-14.

Groark, C. J., & McCall, R. B. (1996). Building successful university-community human service agency collaborations. In C. D. Fisher, J. P. Murray, & E. E. Sigel (Eds.), *Applied developmental science: Graduate training for diverse disciplines and educational settings* (pp. 237-251). Norwood, NJ: Ablex.

Groark, C. J., & McCall, R. B. (2005). Integrating developmental scholarship into practice and policy. In M. H. Bornstein & M. E. Lamb (Eds.), *Developmental psychology: An advanced textbook, 5th Edition,* (pp. 570-601). Mahwah, NJ: Lawrence Erlbaum Associates.

Groark, C. J., & McCall, R. B. (2008). Community-based interventions and services. In M. Rutter, D. Bishop, D. Pine, S. Scott, J. Stevenson, E. Taylor, & A Thapar (Eds.), *Rutter's child and adolescent psychiatry, 5th edition* (pp. 971-988). London, GB: Blackwell Publishing.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika, 87*, 706-710.

Jo, B. (2002). Statistical power in randomized intervention studies with noncompliance. *Psychological Methods*, 7(2), 178-193.

Joyner, E. T., Comer, J. P., & Ben-Avie, M. (Eds.) (2004). *Transforming school leadership and management to support student learning and development.* Thousand Oaks, CA: Corwin Press.

W. K. Kellogg Foundation (1998). *W. K. Kellogg Foundation evaluation handbook*. Battle Creek, MI: W. K. Kellogg Foundation.

W. K. Kellogg Foundation (2000). *W. K. Kellogg Foundation logic model development guide*. Battle Creek, MI: W. K. Kellogg Foundation.

Ludwig, J., & Phillips. D. (2007). The benefits and costs of Head Start. *SRCD Social Policy Report*, *21*(3).

McAllister, C. (1993). The impact of the CCDP on communities in CCDP service areas: Family Foundation's Comprehensive Child Development Program, Ethnographer's Report #10. Washington, DC: Administration for Children, Youth, and Families, Department of Health and Human Services.

McCall, R. B. (in press). Evidence-based programming in the context of practice and policy. *SRCD Social Policy Report*.

McCall, R. B. (1996). The concept and practice of education, research, and public service in university psychology departments. *American Psychologist*, *51*(4), 379-388.

McCall, R. B., & Green, B. L. (2004). Beyond the methodological gold standards of behavioral research: Considerations for practice and policy. *SRCD Social Policy Report, 18*(2), 3-19.

McCall, R. B., Groark, C. J., & Nelkin, R. (2004). Integrating developmental scholarship and society: From dissemination and accountability to evidence-based programming and policies. *Merrill-Palmer Quarterly*, *50*, 326-340.

McCall, R. B., Larsen, L., & Ingram, A. (2003). The science and policy of early childhood education and family services. In A. J. Reynolds, M. C. Wang, & H. J. Walberg (Eds.), *Early childhood programs for a new century: Issues in children's and families lives* (pp. 255-298). The University of Illinois at Chicago Series on Children and Youth. Washington, DC: CWLA Press.

McCall, R. B., Ryan, C. S., & Green, B. L. (1999). Some non-randomized constructed comparison groups for evaluating early age-related outcomes of intervention programs. *American Journal of Evaluation*, *2*(20), 213-226.

Millsap, M. A., Chase, A., Obeidallah, D., Perez-Smith, A., Brigham, N., & Johnston, K. (2000). *Evaluation of Detroit's Comer schools and families initiative*. Cambridge, MA: Abt Associates.

Muthén, L. K., & Muthén, B. O. (1998-2001). *Mplus user's guide*. Los Angeles: Muthén & Muthén.

Olds, D. L., Henderson, C. R. Jr., Kitzman, H. J., Eckenrode, J. J., Cote, R. E., & Tatelbaum, R. C. (1999). In R. E. Behrman (Ed.) *Home visiting: Recent program evaluations. The Future of Children, 61*, 44-65.

Olds, D. L., & Kitzman, H. (1993). Review of research on home visiting for pregnant women and parents of young children. In R. E. Behrman (Ed.), *Home visiting: The future of children, 3,* 53-92. Los Angeles, CA: The David and Lucile Packard Foundation.

Reichardt, C. S., & Rallis, S. F. (Eds.). (1994, Spring). The qualitative-quantitative debate: New perspectives. *New Directions for Program Evaluation Monograph* (No. 61). San Francisco, CA: Jossey-Bass.

Rosenbaum, P. R., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*, 41-55.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach (7th edition)*. Thousand Oaks, CA: Sage.

Schorr, L. B. (2003, February). Determining "what works" in social programs and policies: Toward a more inclusive knowledge base. Washington, DC: The Brookings Institution. Accessed August 20, 2008 at http://www.brookings.edu/papers/2003/0226poverty_schorr.aspx.

Scott, K. G., Mason, C. A., & Chapman, D. A. (1999). The use of epidemiological methodology as a means of influencing public policy. *Child Development, 70*, 1263-1272.

The St. Petersburg-USA Orphanage Research Team (2008). The effects of early social-emotional and relationship experience on the development of young orphanage children. *Monographs of the Society for Research in Child Development,* Serial No. 291, Volume 73 (No. 3).

Wandersman, A., Imm, P., Chinman, M., & Kaftarian, S. (1999*). Getting to Outcomes: Methods and tools for planning, evaluation and accountability*. Rockville, MD: Center for Substance Abuse Prevention.

Wandersman, A., Imm, P., Chinman, M., & Kaftarian, S. (2000*).* Getting to Outcomes: A results-based approach to accountability. *Evaluation and Program Planning*, *23*, 389-395.

Weiss, C. H. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In J. P. Connell, A. C. Kubish, L. B. Schorr, & C. H. Weiss (Eds.), *New approaches to evaluating community initiatives: Concepts, methods, and contexts.* Washington, DC: The Aspen Institute.

Yau, L. H. Y., & Little, R. J. A. (2001). Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with applications to a job training assessment for the unemployed. *Journal of the American Statistical Association, 96*, 1232-1244.

Bibliography

General Applied Issues in Practice and Policy

Groark, C. J., & McCall, R. B. (2005). Integrating developmental scholarship into practice and policy. In M. H. Bornstein and M.E. Lamb (Eds.), Developmental psychology: An advanced textbook, 5th Edition (pp. 570-601). Mahwah, NJ: Lawrence Erlbaum Associates.

Intervention Implementation

Groark, C. J., & McCall, R. B. (2008). Community-based interventions and services. In M. Rutter et al. (Eds.), Rutter's child and adolescent psychiatry, 5th edition. London: Blackwell Publishing.

Gold Standard Methodology

McCall, R. B., & Green, B. L. (2004). Beyond the methodological gold standards of behavioral research: Considerations for practice and policy. SRCD Social Policy Report, 18(2), 3-19.

Program Evaluation and Quasi-Experimental Designs

Cook, T., & Campbell, D. T. (1979). Quasi-experimental design and analysis issues for field settings. Chicago, IL: Rand McNally.

McCall, R. B., Ryan, C. S., & Green, B. L. (1999). Some non-randomized constructed comparison groups for evaluating early age-related outcomes of intervention programs. American Journal of Evaluation, *2*(20), 213-226.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). Evaluation: A systematic approach (7th edition). Thousand Oaks, CA: Sage.

General Longitudinal Methods

Collins, L. M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model. Annual Review of Psychology, 57, 505-528.

McCartney, K., Burchinal, M. R., Rub, K. L. (2006). Best practices in quantitative methods for developmentalists. Monograph of the Society for Research in Child Development, 71(3), No. 385.

Growth Curve Analysis

Burchinal, M. R., & Appelbaum, M. I. (1991). Estimating individual developmental functions: Various methods and their assumptions. Child Development, 62, 23-43.

McCartney et al. (2006). Above, Chapter IV.

Muthén, B. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class-latent profile growth modeling. In L. Collins & A. Sayer (Eds.), New methods for the analysis of change (pp. 289-322). Washington, DC: American Psychological Association.

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), Handbook of quantitative methodology for the social sciences (345-368). Newbury Park, CA: Sage.

Muthén, B. O., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. Psychological Methods, 2, 371-402.

Muthén, B., & Muthén, L. (2000). Integrating person-centered and variable-centered analysis: Growth mixture modeling with latent trajectory classes. Alcoholism: Clinical and Experimental Research, 24, 882-891.

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (2nd ed.). Newburg Park, CA: Sage.

Propensity Score Analysis

Rosenbaum, P. R., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70, 41-55.

Complier Average Causal Effect (CACE) vs. Intent to Treat

Frangakis, C. E., & Rubin, D. B. (1999). Addressing compilations of intention-to-treat analysis in the combined presence of all-or-none treatment-compliance and subsequent missing outcomes. Biometrika, 86, 365-379.

Jo, B. (2002). Statistical power in randomized intervention studies with noncompliance. Psychological Methods, 7(2), 178-193.

Muthén, L. K., & Muthén, B. O. (1998-2001). Mplus user's guide. Los Angeles: Muthén & Muthén.

Yau, L. H. Y., & Little, R. J. A. (2001). Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with applications to a job training assessment for the unemployed. Journal of the American Statistical Association, 96, 1232-1244.

Instrumental Variable Analysis

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. Journal of the American Statistical Association, 91, 444-455.

Davidson, R., & MacKinnon, J. G. (1993). Estimation and inference in econometrics. New York: Oxford University Press.

Foster, E. M., & McLanahan, S. (1996). An illustration of the use of instrumental variables: Do neighborhood conditions affect a young person's chance of finishing high school. Psychological Methods, 1, 249-260.

Yoshikawa, H., Rosman, E. A., & Hsueh, J. (2001). Variation in teen mothers' experience of child care and other components of welfare reform: Selection processes and developmental consequences. Child Development, 72, 299-317.

Policy Relevant Statistics

Scott, K. G., Mason, C. A., & Chapman, D. A. (1999). The use of epidemiological methodology as a means of influencing public policy. Child Development, 70, 1263-1272.