A Research-Service Model for Support of Handicapped Children

MICHAEL J. GURALNICK

Abstract: A model is described in which service programs for handicapped children can be transformed into self contained research-service units. It is suggested that both research and service goals can be achieved concurrently and more economically within this model and that the research component is likely to enhance the effectiveness of the service component as well. The essential characteristics of this approach consist of an intensive individual analysis of each child's behavior and the application of multiple baseline procedures to most aspects of program development.

***Michael J. Guralnick** is Director of Research, National Children's Center, Washington, D.C.*
A version of this article was presented at the 98th Annual Meeting of the American Association on Mental Deficiency, Minneapolis, Minnesota, May 1972.

Agencies at all levels that deal with the support and development of handicapped children are continuously faced with the problems of allocating their limited resources to the various supportive units, such as social work, research, demonstration, or education. These decisions are especially difficult and far reaching in their effects when the units are functionally distinct or where an essential and basic incompatibility exists.

Perhaps the most common conflict arises between areas which may be broadly characterized as belonging to either a research or a service component. This generally results from the fact that the questions posed by research problems normally require carefully controlled situations in order to permit cause and effect statements to be asserted with any degree of confidence. However, when the delivery of services is the primary goal, controlled situations often interfere with the efficient performance of that function. In fact, there are many instances in which an experimental design calls for the exclusion of services for a particular population or segment of that population.

When establishing priorities, the need for services for children requires primary consideration. The vast numbers of children who could benefit from such services, the wide variety of problems that one is faced with, and the continuing battle for support for these services seem sufficient reasons to justify all of one's energies being devoted to purely service aspects. Nevertheless, this author feels that a strong research component cannot only coexist with the service component but can be beneficial and perhaps even necessary for the efficient programing and effective delivery of those services. However, certain adjustments in the type of research, including experimental designs and direction of that research, are required so that the service component is not interfered with. It is the major purpose here to describe

certain procedures in which both goals, research and service, can be achieved concurrently.

This discussion will be restricted to educational intervention programs for preschool youngsters. The comments will be especially appropriate for the many preschool projects, similar in form and function to Head Start, which have been organized for handicapped children and provide programs of a diverse nature. However, the methods described here can be readily applied to a wide range of situations.

Discovering the Critical Components of Successful Programs

Spicker (1971), in a review of procedures and results of various early education intervention programs, noted that, although it is clear that many types of intervention programs can produce intellectual gains as measured by IQ scores, it does not appear that one specific approach is consistently more successful than another. Nevertheless, Spicker seems to have detected some elements that successful programs share in common. He noted that an experimental structure, extensive involvement by a research staff, a positive staff attitude, as well as detailed and daily planning of the curriculum, are factors that appear to be necessary for success. These characteristics, in conjunction with a cognitively or academically oriented program, should, according to Spicker's analysis, produce the greatest gains.

Nevertheless, there are a number of serious problems associated with this type of approach to evaluation in which various groups of children in different types of programs are compared with each other as well as compared with "no program" control groups. Perhaps the most crucial problem is that evaluators and program planners have generally asked of various programs and program centers simply, "Which program is most effective as indicated by our outcome measures?" This question is only a first approximation to a complex analysis and, if not viewed as such, can be misleading.

The significant question, as pointed out by Light and Smith (1970) in referring to Head Start Programs, is, "Which of the program centers worked well for reasons which are known to us and which can be reestablished in any future program centers [p. 12]?" A gross analysis of the outcome variables does not permit statements that answer this second question since each of the centers that may have been evaluated is composed of many different program components, and a systematic assignment of various components to each group generally has not taken place. Since the various components of each program are invariably confounded with each other, it is not possible to determine whether it was, for example, class size, the adult to child ratio, certain aspects of the curriculum, or combinations of these and other factors that were responsible for the observed results. Therefore, researchers are forced to rely on their intuition and tenuous post-hoc analyses and to engage in speculation with respect to their statements about why a program was so effective.

Light and Smith (1970) further criticized these designs, pointing out that, in those methods which employ retrospective matching to obtain equivalence of control and experimental groups, an adequate basis for comparisons of the various groups is rarely obtained since matching is imperfect, even when used in conjunction with analyses of covariance. More importantly, they noted that in designs in which children were randomly assigned to various treatment and control groups the analysis of the results usually takes the form of assessing the average gains of the treatment in comparison to controls or other treatments. They argued that the variability within a treatment, as well as the average, is equally important, pointing out that a sizeable percentage of children in a program can actually fare more poorly than controls while the average score suggests the existence of an effective program. In effect, the point is often brought home that the crucial question is what specific component or components of a program are effective, and it may be that they are effective for only a limited segment of the treatment population.

Light and Smith offered an alternative approach, referred to as an exploratory-experimental model in which potentially effective and externally controllable components are identified and then systematically introduced into succeeding generations of programs. These controllable features are then isolated and evaluated across a number of programs, and the successful aspects are combined in an attempt to produce an even better program. This exploratory model is an important concept and should receive considerable attention in the future. However, this may not be a totally adequate approach since the emphasis on overall outcome measures rather than the process of learning still remains, as well as costs associated

with the sizeable number of programs and long term evaluations which would be needed to isolate the effective components.

Individual Analysis

From a completely different perspective, Sidman (1960) argued that an intensive analysis of each individual's behavior is the most appropriate approach to the understanding of behavior. Individual differences in performance in response to a treatment simply present a challenge for the experimenter to identify those sources of variability. Perhaps the most important characteristic for this discussion is the emphasis on the process of learning rather than the product.

This approach required a new set of experimental designs that have now been applied by laboratory psychologists studying operant behavior and those applying reinforcement principles to various practical situations. In all instances, the purpose is to specify as precisely as possible the factors which control behavior. Briefly, operant conditioning methodology typically uses two classes of experimental designs to demonstrate that a given variable is in fact the controlling variable (Baer, Wolf, & Risley, 1968). By *variable* it is meant any specifiable set of procedures that can be reliably administered to individuals. These might include things such as segments of an instructional unit for number concepts or a method to reduce disruptive behavior by time-out from reinforcement.

The first design is the familiar reversal procedure in which baseline conditions and treatment conditions are alternated to demonstrate control. Although this can be a powerful experimental procedure, there are many instances in which it cannot be employed (see Bandura, 1969). In these cases the second class of experimental procedures, referred to as multiple baseline designs, is appropriate. In this procedure, a number of baselines of several behaviors from one or several individuals are obtained simultaneously, and the experimental treatment is applied successively to each separate behavior. If the treatment variable is truly effective, then any specific behavior will change accordingly when the variable is applied but will remain relatively stable until that time. Each demonstration of this variable's effect increases its reliability as being the variable that has, in fact, produced the change.

This type of design has been recently elaborated upon by Hall, Cristler, Cranston, and Tucker (1970), and many variations directly applicable to the classroom setting have been described. In addition, multiple baseline procedures have been used to evaluate the effects of programs designed to reduce various undesirable behaviors (Barton, Guess, Garcia, & Baer, 1970; Schwarz & Hawkins, 1970), as well as for programs concerned with the development of appropriate behavior, especially language (Schumaker & Sherman, 1970; Twardosz & Sajwaj, 1971).

Description of Research-Service Model

The behavioral techniques just discussed are generally used to assess the efficacy of reinforcement variables such as the quality, amount, and schedule of reinforcements as they affect a wide range of behaviors. However, the crucial point for this discussion is that multiple baseline procedures can be adapted to assess the efficacy of instructional programs as well. Accordingly, an alternative to the Light and Smith model can be suggested, although the purposes, i.e., identifying the controllable, specifiable, and effective aspects of programs, are identical.

The alternative recommended here is to turn each service program into a self contained research-service unit and to conduct programmatic research while simultaneously proving educational and other supportive services. In doing so, between-group comparisons are dispensed with, and attention is focused directly on a detailed analysis of the environmental effects on each individual child's behavior while multiple baseline procedures are applied to establish cause and effect relationships. Adoption of this model should considerably reduce the problems associated with the guesswork involved in identifying the effective variables. From the point of view of economy, many of the procedures can be adequately carried out by the teaching staff functioning as teachers and researchers (Hall et al., 1970) and can be accomplished without significantly affecting their teaching effectiveness. Moreover, it avoids the additional expense, both financial and educational, that results when programs fail or when information about apparently successful programs is obscure or imprecise.

Procedures

The technique to accomplish this goal of conducting a combined research-service program requires that each instructional program,

e.g., to teach number concepts, alphabet discrimination, or use of a specific verbal response form, first be carefully broken down into specifiable units, with each unit containing both a detailed procedure and a clearly defined behavioral goal. Once this is accomplished, baseline behaviors for the unit for several individuals are recorded.

Following this, the instructional unit is administered to a child. Behavioral criteria will have been previously established to form the basis for an assessment of the value of the instructional unit. These factors should include, among others, the rate of response, the number and pattern of errors, the amount of time or number of lessons required to reach criterion, the extent of generalization, and the degree of teacher involvement that was necessary. It should be noted that there are no hard and fast rules to indicate what numbers should be attached to these criteria. This will depend on a number of factors, especially other work dealing with this content area and previous experiences of the teachers.

If this instructional unit is rejected as either being totally ineffective or is judged inefficient on various grounds, a revision is constructed and again administered. If the child now responds appropriately, the revised version is administered to another child. If the baseline behavior of the second child had not changed substantially before but now changes appropriately with the presentation of the revised program, then confidence in the efficacy of the materials is increased. These same procedures are then applied to a third child for a further assessment and should include administering the original program prior to the revised version (anticipating little change at this point) as a check on the possible effect of revisions per se.

The end product of this process is a clear statement that this particular instructional unit does work, that is, it meets certain behavioral criteria that may include efficiency, accuracy, or other time and context related factors. Note that the procedure has been so devised that it is possible to rule out extraneous factors such as the passage of time, general attention to target behaviors, or previous revisions of the instructional programs as possible causal agents. As a consequence, statements can be made about causality, and these are, of course, answers to research questions. Variables such as group size, nature of reinforcement, extent of prompting, and all of the various interactions can be accommodated by this procedure.

Variability

Problems in interpretation will arise if changes occur immediately in some instances and more slowly in others or if there is substantial intersubject variability. The important point here is that effective instructional materials, designed even for a restricted population, should have sufficient flexibility built into the materials to accommodate a reasonably wide range of individual differences. Consequently, if such variability does result, revisions of the instructional materials will be required. In doing so, variability is not, as Sidman noted, "buried in the standard deviation" as occurs in group analyses but rather occasions the exploration of the variables that contributed to the variability. With regard to this, Bricker (1970) noted:

> [I]f a program is 80 percent replicable for a given population, then we must determine why the remaining 20 percent are not learning. Although something may be wrong with the nervous systems of these nonlearners, we are better advised to modify the program being used, or to devise a new program that will suit the learning needs of this 20 percent. An analysis of the frame-by-frame performance of this 20 percent can yield important indicators as to where and in what general ways the program fails [p. 20].

Ideally, a careful analysis of the characteristics of children who show must variability across instructional programs or reinforcement procedures may reveal certain interesting and consistent correlations with historical variables that, in turn, can serve as warning signals for future attempts at instruction or can be used for various theoretical purposes. As such, this is consistent with interactive models (Reynolds & Balow, 1972) in which the differential effects of treatments are studied in relation to sociological, aptitude, or other historical characteristics of individuals. In effect, an instructional procedure may be effective for the overall population, but it may turn out that a substantial proportion of individuals for whom the treatment was totally ineffective may share some common characteristic such as being below some point on the socioeconomic scale. Additional efforts here may reveal that it is possible to identify various subgroups of individuals and to devise a different behavioral procedure for each subgroup. This may have the consequence of abolishing substantial individual deviations (Owens, 1968).

It should also be pointed out that certain aspects of the multiple baseline procedure described here frequently constitute a systematic replication of the effectiveness of the instructional procedures. In the explicit application of this procedure, various parameters under which the original behavioral procedure was established are varied to determine if the result can still be obtained. If the behavior remains invariant with these changes, an unusually powerful demonstration of the reliability and generality of these methods has been provided (Sidman, 1960).

Since, in practice, a particular instructional unit is presented to different children in different situations at different times, after the necessary revisions have been made, there exists a situation that approximates the conditions for systematically replicating the results. This can be made more explicit and systematic by, for example, asking different teachers to administer the instructional materials. If the outcome is unchanged despite this variation, confidence in the effectiveness of this instructional unit is enhanced, especially since past experience has revealed the teacher to be a potent variable in many other situations.

On the other hand, if the instructional materials are not as effective with the introduction of this variable, direct replications of the original findings are needed and then an intensive investigation of the interaction between teacher characteristics and the behavior under study. Accordingly, this approach represents appreciation of the dynamic and interactive nature of the complex social situation called learning, yet it provides a technology capable of dealing with the problems and issues on that very level.

The problem remains, however, as to how much variability can be tolerated. This question cannot be answered in absolute terms since, like questions concerning the meaning and reliability of statistical analyses, the problem goes beyond purely experimental and methodological considerations in that the practical aspects of the situation must receive considerable weight. With respect to the practical aspects it must be decided how important it is that the specific unit be almost unquestionably established as the controlling variable. If an error has been made in accepting a particular unit as generally effective, can adjustments be made at a later time? Moreover, questions about the extent to which this will affect the overall effectiveness of the instructional program, as well as the percentage of the population that is likely to find difficulty with the program, are appropriate here.

Limitations and Variations

This approach can only answer questions pertaining to the programmatic aspects of an intervention program. As such, there are many variables, especially those relating to the administration of programs and long term goals, that are important but are not testable within this model. Also, there may be certain instances in which a small scale between-groups experiment, within the program, to answer a highly specific question is appropriate. However, if this occurs it should never take the form of an experimental-control comparison but rather a comparison of, for example, two techniques, both likely to produce positive results but which require different programing approaches. In some instances, the multiple-baseline procedure can be employed within each group, thereby providing the basis for an individual analysis as well. Finally, certain problems may permit the application of various complex experimental designs, such as a Latin square design for individual subjects (Browning, 1967), but only limited information can be gained here.

Effect of Research on the Service Component

An important question concerns the effect of the research orientation on the service component. Although some time and effort must be given over to the research aspects of the program, available evidence indicates that it is likely to be more than balanced out by the positive effects resulting from the experimental approach. As noted earlier, Spicker (1971) emphasized that probable key features, among others, of successful early childhood intervention programs are that they are highly structured, are well supervised, and have access to a research staff. Weikert (1972) reached a similar conclusion, stressing the critical importance of detailed and daily operational planning and the necessity of supervision. These are, of course, requirements in the joint research-service program proposed here.

Moreover, this program focuses on the process of learning, rather than the product; the effects of instructional inputs upon each discriminable unit of the procedure are carefully examined. In effect, the relative effectiveness of each of the various components of any of the instructional units can be analyzed.

Finally, these methods can, of course, be directly applied to assess techniques designed to influence noncognitive variables such as the "survival skills" identified by Cobb (1970). The explicit development and experimental analysis of these skills are usually lacking when gross measures of outcome are the focus of attention.

Concluding Comments

There is a close correspondence between the methods described here and those of precision teaching (see *TEACHING Exceptional Children,* 1971, Vol. 3, No. 3). However, this model takes one further step by asking that the instructional programs and reinforcement procedures be administered in a systematic fashion so as to permit a precise evaluation of those activities. In doing so, services for handicapped children can be provided at the same time the instructional methods are being experimentally evaluated. The end product should be a more efficient and effective set of procedures and methods that benefit both components.

References

Baer, D. M., Wolf, M. M., & Risley, T. R. Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1968, **1**, 91-97.

Bandura, A. *Principles of behavior modification.* New York: Holt, Rinehart, & Winston, 1969.

Barton, E. S., Guess, D., Garcia, E., & Baer, D. M. Improvement of retardates' mealtime behaviors by timeout procedures using multiple baseline techniques. *Journal of Applied Behavior Analysis,* 1970, **3**, 77-84.

Bricker, W. A. Identifying and modifying behavioral deficits. *American Journal of Mental Deficiency*, 1970, **75**, 11-21.

Browning, R. A same-subject design for simultaneous comparison of three reinforcement contingencies. *Behavior Research and Therapy*, 1967, **5**, 237-243.

Cobb, J. A. Survival skills and first grade academic achievement. Report No. 1. *University of Oregon Center for Research and Demonstration in The Early Education of Handicapped Children.* Report to the Office of Education, Washington, D.C., 1970.

Hall, R. V., Cristler, C., Cranston, S. S., & Tucker, B. Teachers and parents as researchers using multiple baseline designs. *Journal of Applied Behavior Analysis*, 1970, **3**, 247-255.

Light, R., & Smith, P. Choosing a future: Strategies for designing and evaluating new programs. *Harvard Educational Review*, 1970, **40**, 1-28.

Owens, W. A. Toward one discipline of scientific psychology. *American Psychologist*, 1968, **23**, 782-785.

Reynolds, M. C., & Balow, B. Categories and variables in special education. *Exceptional Children,* 1972, **38**, 357-366.

Schwarz, M. L., & Hawkins, R. P. Application of delayed reinforcement procedures to the behavior of an elementary school child. *Journal of Applied Behavior Analysis*, 1970, **3**, 85-96.

Schumaker, J., & Sherman, J. A. Training generative verb usage by imitation and reinforcement procedures. *Journal of Applied Behavior Analysis*, 1970, **3**, 273-287.

Sidman, M. *Tactics of scientific research.* New York: Basic Books, 1960.

Spicker, H. S. Intellectual development through early childhood education. *Exceptional Children*, 1971, **37**, 629-640.

Twardosz, S., & Sajwaj, T. Development of complex sentence structure in a retarded girl using a multiple base-line technique. *Proceedings of the 79th Annual Convention of the American Psychological Association*, 1971, **6**, 663-664.

Weikert, D. P. Relationship of curriculum, teaching, and learning in preschool education. In J. C. Stanley (Ed.), *Preschool programs for the disadvantaged: Five experimental approaches to early childhood education.* Baltimore: Johns Hopkins University Press, 1972. Pp. 22-66.