
Walter Williams
John W. Evans

A far-reaching controversy has flared over a recent Westinghouse Learning Corporation-Ohio University evaluation study showing that Head Start children now in the first, second, and third grades differed little, on a series of academic achievement and attitudinal measures, from comparable children who did not attend Head Start.


The Politics of Evaluation

In the heat of the public controversy, there have been some old-fashioned political innuendos based on vile motives, but, in the main, the principal weapons in the battle have been the esoteric paraphernalia of modern statistical analysis. This is appropriate; the methodological validity of the Head Start study is a critical part of the debate. However, the real battle is not over the methodological purity of this particular study, but, rather, involves fundamental issues of how the federal government will develop large-scale programs and evaluate their results.

At this deeper level of the debate, what we are seeing is a head-on collision between two sets of ideas developed in the mid-1960's. On the one hand, there was the implicit premise of the early years of the war on poverty that effective programs could be launched full-scale, and could yield significant improvements in the lives of the poor. Head Start was the archetype of this hope. Born in late 1964, the program was serving over a half-million children by the end of the following summer. On the other hand, the federal government, during roughly the same period, implemented the Planning, Programming, Budgeting System (PPBS), founded on the premise that rigorous analysis could produce a flow of information that would greatly improve the basis for decision-making. And the notion of evaluating both ongoing programs and new program ideas was fundamental to this type of thinking.

To see the dimensions and ramifications of this clash, it is necessary to return to those halcyon days in which the basic ideas of the war on poverty and PPBS were formulated. Only then can we explore the present Head Start controversy to see what we may learn from it for the future.

THE EARLY DAYS OF THE WAR ON POVERTY

On June 4, 1965, President Johnson said in his Howard University Address, entitled "To Fulfill These Rights":

To move beyond opportunity to achievement ... I pledge you tonight this will be a chief goal of my administration, and of my program next year, and in years to come. And I hope, and I pray, and I believe, it
Policy Evaluation

will be a part of the program of all America. ... It is the glorious
opportunity of this generation to end the one huge wrong of the
American Nation and, in so doing, to find America for ourselves, with
the same immense thrill of discovery, which gripped those who first
began to realize that here, at last, was a home for freedom.

The speech rang with hope—a call for basic changes that seemed
well within our grasp. Viewed from the present, the address marked
a distinct watershed. It was the crest of our domestic tranquillity,
based on the strong belief that black and white could work together
in harmony as a nation. The speech also marked the high point of our
faith in our ability to bring about significant change. Despite some of
the rhetoric of the time to the effect that change would not be easy,
it is fair to say that the faith was there that giant steps could be taken
quickly. On that June day, there was the strong belief that the concen-
trated effort of the war on poverty, launched less than a year before,
could bind the nation together.

This faith had two dimensions—first, that there could be a redistrib-
ution of funds and power toward the disadvantaged and, second,
that, with such a redistribution, new programs could bring substantial
improvement in the lot of the disadvantaged. The first was both more
clearly perceived and more glamorous. To wrest power and money
from the entrenched forces was heady stuff. Less clearly perceived was
that redistribution was a necessary, but not a sufficient, condition of
progress. New programs had to be devised, not just in broad brush
strokes, but in the nitty-gritty detail of techniques and organization.

Taking young black men from the ghettos to the wilderness of an
isolated Job Corps Center was not a solution in itself. One had to
worry about such mundane things as curriculum and the morale of
these young men in a Spartan, female-absent environment. This atmo-
sphere of confidence and enthusiasm led us to push aside the fact that
we had neither the benefit of experience in such programs nor much
realization of the difficulties involved in developing effective techniques.

Standing in 1969 on the battle-scared ground of the war on pov-
erty, it is easy to see the naiveté and innocence of that time—scarcely
half a decade ago. Events were to crash upon us quickly. Vietnam was
to end any hope for large funds. Riots, militancy, and the rise of
separatism made the earlier ideas of harmony seem quaint. Those with
established power did not yield easily either to moral suasion or to

more forceful means. Real power is still a well-guarded commodity.

Most important for this discussion, we have found, over a wide
range of social-action programs, both how unyielding the causes of
poverty are and how little we really know about workable techniques
for helping the disadvantaged. The point is not that we are unable to
derive "reasonable" programs from bits and pieces of information and
hard thinking. We can, we have. But our experience seems to point
up, over and over again, the almost insurmountable difficulty of bridg-
ing the gap between brilliantly conceived programs and those which
work in the field. Great pressures exist for new "solutions" to social
problems to be rushed into national implementation as soon as they
are conceived. But the attempts to go directly from sound ideas to
full-scale programs seem so often to end in frustration and disappoint-
ment.

THE ORIGINS OF ANALYSIS WITHIN THE GOVERNMENT

In the early 1960's, Secretary Robert McNamara relied on a conцеп-
tual framework, formulated at the RAND Corporation, to make
analysis a critical factor in the decision-making process of the Depart-
ment of Defense. In October 1965, drawing on this experience, the
Bureau of the Budget issued Bulletin No. 66-3, establishing the Plan-
ning, Programming, Budgeting System within all federal departments
and agencies. The departments and agencies were instructed to "es-
establish an adequate central staff or staffs for analysis, planning, and
programming [with] ... the head of the central analytical staff ... 
directly responsible to the head of the agency or his deputy." These
central offices were to be interposed between the head of the agency
and the operating programs and were charged with undertaking anal-
ysis that would provide a hard quantitative basis on which to make
decisions. For social-action agencies, this was a radical change in the
way of doing business.

Before PPBS, not much progress had been made in analyzing
social-action programs. Although the broad approach developed at
the Department of Defense might be used in such analyses, the rele-
vance of particular methodological tools was less clear. For example,
there was little actual experience with the kinds of evaluations which
seek to measure the effects of a social-action program on its partici-
pants or the external world. And a host of formidable problems
Policy Evaluation

existed, such as the lack of good operational definitions for key variables, the shortage of adequate test instruments, and the difficulties of developing valid control groups. Thus, the usefulness of evaluative analysis for social-action programs would have to be proved in particular situations.

Beyond this was the political question of bringing analysis into the agency's policy-making process. As analytical studies were quite new to social-action programs, their results—especially those measuring the effectiveness of ongoing programs—were seen as a threat by those with established decision-making positions. Unfavorable evaluation results have a potential either to restrict a program's funds or to force major changes in the direction of the program. One can hardly assume passive acceptance of such an outcome by the managers and operators of programs.

Thus, one can see how the tiny dark cloud of the Head Start controversy formed at this early date. The push toward new operating programs and the emerging PPBS brought about a role conflict between those who ran programs (and believed in them) and those who analyzed these programs (and whose job it was to be skeptical of them). As former Director of the Bureau of the Budget Charles L. Schultze has observed:

"In the relationship between the political process and the decision-making process as envisaged by PPB... I do not believe that there is an irreconcilable conflict... But they are different kinds of systems representing different ways of arriving at decisions. The two systems are so closely interrelated that PPB and its associated analytic method can be an effective tool for aiding decisions only when its relationships with the political process have been carefully articulated and the appropriate roles of each defined... It may, indeed, be necessary to guard against the naiveté of the systems analyst who ignores political constraints and believes that efficiency alone produces virtue. But it is equally necessary to guard against the naiveté of the decision maker who ignores resource constraints and believes that virtue alone produces efficiency."


The Politics of Evaluation

Looking in retrospect, at the early PPBS vis-à-vis social-action programs, it may be said that: (1) the absolute power of analysis was oversold and (2) the conflicts in the system between the analytical staff and the operators of the programs was underestimated. Hence, the politics of evaluation—in essence, the clash between methodology, political forces, and bureaucracy—looms much larger than was imagined in those early days. At the same time, knowing more today about how difficult it is to develop and operate effective programs, the need for analysis—the need to assess both our current operations and our new ideas—seems even more pressing than in the less troubled days of 1965.

BACKGROUND OF THE HEAD START STUDY

With these general considerations as background, we now need to look briefly at the key elements within OEO: the Head Start program; OEO's analytical office, the Office of Research, Plans, Programs, and Evaluation (RPP&E); and the general state of evaluation of the antipoverty programs prior to the Westinghouse study.

Head Start

The concepts underlying Head Start were based on the thinking of some of the best people in the child-development area and on a variety of research findings (probably relatively rich compared to most other new programs) suggesting a real potential for early childhood training, but offering few and often conflicting guidelines as to the detailed types of programs to be developed. In fact, the original concept of Head Start was that it was to be an explicitly experimental program reaching a limited number of children. The idea, however, was too good. It was an ideal symbol for the new war on poverty. It generated immediate national support and produced few political opponents. In this atmosphere, one decision led easily to another, and Head Start was quickly expanded to a $100 million national program serving a half-million children. In the beginning, Head Start consisted mainly of six-to-eight-week summer projects under a variety of sponsors (school systems, churches, and community-action agencies, for example) with a high degree of local autonomy concerning how the project
was to be carried out. Later, Head Start funded a significant number of full-year projects with a similar policy of flexibility and local autonomy.

The immense popularity of the early days carried over. Head Start remained OEO's showcase program, supported strongly by the Congress, communities, poor mothers, and a deeply committed band of educators (many with a significant personal involvement in the program).

RPP&E

Analysis came early to OEO because its Office of Research, Plans, Programs, and Evaluation was one of the original independent staff offices reporting directly to the head of the agency. RPP&E predated the PPBS Bulletin, but was, in many ways, the epitome of the PPBS analytical staff, in that it was headed by RAND alumni who stressed the power of analysis. RPP&E was both a major developer of analytical data and a key factor in the agency's decision-making process. As one might expect, in this role it had more than once clashed with program-operators.

Evaluation at OEO

Critical to our discussion is the fact that RPP&E did not establish a separate Evaluation Division until the autumn of 1967. Prior to that time, most of the responsibility for evaluation rested with the programs, but RPP&E had had some involvement, particularly in trying to use data developed by the programs to make over-all program assessments.

In the case of Head Start, the program itself had initiated a large number of individual project-evaluations, mainly of the summer program. Across a wide range of these projects it was found that, in general, participants who had been given various cognitive and affective tests at the beginning of the Head Start program showed gains when tested again at the end of the program. However, virtually all the follow-up studies found that any differences which had been observed between the Head Start and control groups immediately after the end of Head Start were largely gone by the end of the first year of school. The meaning of this "catch up" by the control group has been and still is subject to considerable debate, ranging from doubts that the immediate post-program gains were anything more than test-retest artifacts, to assertions that the superior Head Start children raise the performance levels of their non-Head Start classmates.

RPP&E had tried fairly early to develop its own national assessments of Head Start, but found little support for such undertakings within the program. Two such studies were developed, but the results were marred by technical and analytical problems. At the time of the establishment of the Evaluation Division, therefore, no good evidence existed as to over-all Head Start effectiveness—a fact that was beginning to concern the agency, the Bureau of the Budget, and some members of Congress.

As one might guess, the program offices hardly greeted the newly created Evaluation Division with enthusiasm—no one was happy with a staff office looking over his shoulder. In a formal division of labor, three types of evaluation were recognized. RPP&E was given primary responsibility for evaluation of the over-all effectiveness of all OEO programs (Type I). The programs retained primary responsibility for both the evaluation of the relative effectiveness of different program strategies and techniques, for example, different curricula in Head Start (Type II) and the on-site monitoring of individual projects (Type III). The basic logic of this division of labor was to ensure that Type I over-all evaluations would be carried out, to locate the responsibility for these evaluations at a staff-office level removed from the programs, and, at the same time, to place the Type II and Type III evaluation-responsibilities at the program level because of the greater need for detailed program-knowledge that these kinds of evaluation require.

This division of labor also matches the type of evaluation with the types of decisions for which different levels within the organization have primary responsibility—the over-all mixture of programs and resource-allocation at the top (Type I), and program design (Type II) and management (Type III) at the program level.

THE WESTINGHOUSE STUDY

Thus, it was out of this total complex of conditions that the Westinghouse evaluation of Head Start originated.
Policy Evaluation

— The explosive expansion of Head Start from what was originally conceived as a limited experimental program to a large national program almost overnight.
— A developing commitment throughout the government to increasing analysis and assessment of all government programs.
— The national popularity of the Head Start program and the widespread equation of this popularity with effectiveness.
— Previous evaluations of Head Start that did not provide adequate information on the program’s over-all impact.
— The development of a new staff-level evaluation function at OEO charged with producing timely and policy-relevant evaluations of the over-all impact of all OEO programs.

As one in a series of national evaluations of the major OEO programs, the new RPP&E Evaluation Division proposed for the Head Start program an ex post facto study design in which former Head Start children, now in the first, second, and third grades of school, were to be tested on a series of cognitive and affective measures, and their scores compared with those of a control group. Because the program was in its third year and there was, as yet, no useful assessment of its over-all effects, time was an important consideration in deciding on an ex post facto design. Such a design would produce results relatively soon (less than a year), as compared with a methodologically more desirable longitudinal study which would take considerably longer.

Within the agency, Head Start administrators opposed the study on a number of grounds, including the inadequacy of the ex post facto design, the weakness of available test instruments, and the failure to include other Head Start goals such as health, nutrition, and community involvement. In sum, Head Start contended that this limited study might yield misleading negative results which could shake the morale of those associated with Head Start and bring unwarranted cutbacks in the program. RPP&E evaluators did not deny the multiplicity of goals, but maintained that cognitive improvement was a primary goal of Head Start and, moreover, was an outcome which reflected, indirectly, the success of certain other activities (for example, better health should facilitate better school performance). Further, the study’s proponents in RPP&E recognized the risks outlined by Head Start officials, but argued that the need for evaluative evidence in order to improve the decision-making process makes it necessary to run these risks. After much internal debate, the Director of OEO ordered that the study should be done, and a contract was made in June 1968 with the Westinghouse Learning Corporation and Ohio University.

The study proceeded in relative quiet, but as it neared completion, hints came out of its negative findings. Because President Nixon was preparing to make a major address on the poverty program, including a discussion of Head Start, the White House inquired about the study and was alerted to the preliminary negative results. In his Economic Opportunity Message to the Congress on February 19, 1969, President Nixon alluded to the study and noted that “the long-term effect of Head Start appears to be extremely weak.”

This teaser caused a flood of requests for a full disclosure of the study’s findings. In the Congress, where hearings were being held on OEO legislation, strong claims were made that OEO was holding back the results to protect Head Start. This was not the case, but the demands did present a real dilemma for the agency—particularly RPP&E. For the results at that time were quite preliminary, and Westinghouse was in the process of performing further analysis and verification of the data. Hence, RPP&E, which, in general, was anxious for evaluative analysis to have an impact at the highest levels of government, did not want to suffer the embarrassment of a national debate over tentative results that might change materially in the later analysis. However, after much pressure, an early, incomplete version of the study was released. In June, the final report was published, and it confirmed the preliminary findings.

These background facts are important in order to understand why the controversy rose to such a crescendo, as it ranged over the executive branch and the Congress, with wide coverage in the press. The Westinghouse study is, perhaps unfortunately, an instructive example of public reaction to evaluations of social-action programs. As we turn now to a brief description of the study, its findings, and a discussion of its methodological and conceptual base, this milieu must be kept in mind.

The study and its major conclusions are summarized succinctly in the following statement by the contractor:
Policy Evaluation

The basic question posed by the study was:

To what extent are the children now in the first, second, and third grades who attended Head Start programs different in their intellectual and social-personal development from comparable children who did not attend?

To answer this question, a sample of one hundred and four Head Start centers across the country was chosen. A sample of children from these centers who had gone on to the first, second, and third grades in local area schools and a matched sample of control children from the same grades and schools who had not attended Head Start were administered a series of tests covering various aspects of cognitive and affective development [The Metropolitan Readiness Test, the Illinois Test of Psycholinguistic Abilities, the Stanford Achievement Test, the Children's Self-Concept Index, and the like.] The parents of both the former Head Start enrollees and the control children were interviewed and a broad range of attitudinal, social, and economic data was collected. Directors or other officials of all the centers were interviewed and information was collected on various characteristics of the current local Head Start programs. The primary grade teachers rated both groups of children on achievement motivation and supplied a description of the intellectual and emotional environment of their elementary schools.

Viewed in broad perspective, the major conclusions of the study are:

1. Summer programs appear to be ineffective in producing any gains in cognitive and affective development that persist into the early elementary grades.
2. Full-year programs appear to be ineffective as measured by the tests of affective development used in the study, but are marginally effective in producing gains in cognitive development that could be detected in grades one, two, and three. Programs appeared to be of greater effectiveness for certain subgroups of centers, notably in mainly Negro centers, in scattered programs in the central cities, and in Southeastern centers.
3. Head Start children, whether from summer or from full-year programs, still appear to be considerably below national norms for the standardized tests of language development and scholastic achievement, while performance on school readiness at grade one approaches the national norm.

The Politics of Evaluation

4. Parents of Head Start enrollees voiced strong approval of the program and its influence on their children. They reported substantial participation in the activities of the centers.

In sum, the Head Start children cannot be said to be appreciably different from their peers in the elementary grades who did not attend Head Start in most aspects of cognitive and affective development measured in this study, with the exception of the slight, but nonetheless significant, superiority of full-year Head Start children on certain measures of cognitive development.

METHODOLOGICAL ISSUES

We now turn to the methodological and conceptual validity of the study—the explicit focal point of the controversy—and this presents difficult problems of exposition. First, both of us are protagonists on one side of the controversy, with Evans being one of the major participants in the debate. Second, a presentation of the methodological questions in sufficient detail to allow the reader to form his own opinions would require an extensive discussion. The final Westinghouse report comprises several hundred pages, with a significant portion of it directed specifically to methodological issues. Under these circumstances, we will summarize the major criticisms that have been made of the study and comment on them briefly in this section. Then, in the next major section, we will set out our judgment as to the over-all technical adequacy of the report and its usefulness for decision-making.

Criticisms of the Study

1. The study is too narrow. It focuses only on cognitive and affective outcomes. Head Start is a much broader program which includes health, nutrition, and community objectives, and any proper evaluation must evaluate it on all these objectives.

Policy Evaluation

Our experience has been that one of the reasons for the failure of so many evaluations is that they have aspired to do too much. We did not think that it was possible to cover all of the Head Start objectives in the same study; therefore, we purposely limited the study's focus to those which we considered most important. Despite its many other objectives, in the final analysis Head Start should be evaluated mainly on the basis of the extent to which it has affected the life-chances of the children involved. In order to achieve such effects, cognitive and motivational changes seem essential.

2. The study fails to give adequate attention to variations among the Head Start programs. It lumps the programs together into an over-all average and does not explore what variation there may be in effectiveness as a function of differing program styles and characteristics. The study, therefore, fails to give any guidance concerning what detailed changes (for example, types of curricula) should be made in the program.

This is essentially correct. As discussed earlier, the purpose of the evaluation was to measure the over-all effectiveness of the Head Start program in a reasonably short period of time. This in no way denies the need for a study to get at the question of variation among the programs. The fact is that both over-all and detailed information are frequently needed, but the latter generally takes much longer to develop.

3. The sample of full-year centers in the study is too small to provide confidence in the study's findings. Because of such a small sample, the lack of statistically significant differences between the Head Start and control groups is to be expected, and gives a misleading indication that the programs had no effect. With such a small sample, it would take quite large differences to reach a satisfactory level of statistical significance.

The 104 Head Start centers, selected at random, were chosen in order to provide an adequate total sample. This was then broken down into an approximate 70-30 division in order to approximate the actual distribution of summer and full-year programs. If we were doing the study over, we would select a larger number of full-year centers. The main advantage, however, would be to allow more analysis of subgroups within the full-year sample. It is very unlikely that the study's principal conclusions about the over-all effectiveness of the program would be altered by a larger sample. A detailed "power of the test" analysis showed that with the present sample size and variance, the statistical tests are capable of detecting differences between the experimental and control groups below the level of what would be practically meaningful. Forgetting the statistical complexities for a minute, the simple fact is that the differences between the Head Start and control-group scores were quite small. Even in the cases in which differences were statistically significant, they were so small as to have little practical importance.

4. The sample is not representative. Many of the original randomly chosen centers had to be eliminated.

The study suffered a loss of some of the centers specified in the original sample because (1) some small rural areas had all their eligible children in the Head Start program (and hence no controls could be found) and (2) some communities prohibited the testing of children in the school system. Centers were substituted randomly, and a comparison of the final chosen sample with the total universe of Head Start centers showed the two to be very similar on a large number of factors (for example, rural-urban location, racial composition, and the like).

5. The test instruments used in this study, and indeed all existing instruments for measuring cognitive and affective states in children, are primitive. They were not developed for disadvantaged populations, and they are probably so gross and insensitive that they are unable to pick up many of the real and important changes that Head Start has produced in children.

It is entirely possible that this is true. However, most of the cognitive measures are the same ones being used by other child-development and Head Start researchers doing work on disadvantaged children. In those cases (relatively few) where previous studies have shown positive changes on these very same measures, they have seldom been questioned or disregarded because of the inadequacy of the instruments. In the affective area, Westinghouse found no appropriate
test instruments and had to devise its own. Hence, the results should be viewed as suggestive, but no more. The Westinghouse study used the best instruments available, and with these instruments, few appreciable differences are found between children who had been part of a Head Start program and those who had not.

6. The study is based on an ex post facto design which is inherently faulty because it attempts to generate a control group by matching former Head Start children with other non-Head Start children. A vast number of factors, either alone or acting together, could produce a superior non-Head Start group which would obscure the effect of the program.

It is always possible in any ex post facto study that failure to achieve adequate matching on all relevant variables (particularly self-selection factors) can occur. Ex post facto studies, however, are a respected and widely used scientific procedure, although one which does not provide the greater certainty which results from the classic before-after experimental design carried out in controlled laboratory conditions.

In the Westinghouse study, the two groups were matched on age, sex, race, and kindergarten attendance. Any residual differences in socioeconomic status were equated by two different statistical procedures: a random-replication-covariance analysis and a nonparametric matching procedure. Both statistical techniques, which equated the two groups on parent’s occupation, education, and per capita income, yielded the same basic results on the cognitive and affective comparisons between Head Start and control-group children.

7. The study tested the children in the first, second, and third grades of elementary school—after they had left Head Start. Its findings merely demonstrate that Head Start achievements do not persist after the children return to poverty homes and ghetto schools. Rather than demonstrating that Head Start does not have appreciable effects, the study merely shows that these effects tend to fade out when the Head Start children return to a poverty environment.

It is possible that poor teachers, impoverished environment, and other similar factors eliminated a significant cognitive advantage gained by Head Start children during the Head Start period. But even if this is true, we must have real doubts about the current course of the program. Unless Head Start alone can be improved so as to have positive effects which do not disappear, or unless Follow Through or some other program can be developed to provide subsequent reinforcement that solidifies the gains of Head Start children, the present worth of the gains seems negligible. Whatever the cause, the fact that the learning gains are transitory is a most compelling fact for determining future policy.

8. The study’s comparison of Head Start with non-Head Start children in the same classrooms fails to take into account secondary or spillover effects from the Head Start children. The children who have had Head Start are likely to infect their non-Head Start peers with their own greater motivation and interest in learning. Their presence in the classroom is also likely to cause the elementary school teacher to upgrade her entire level of teaching or to give more attention to, and therefore produce greater gains in, the less advanced non-Head Start group. Thus, the study minimizes Head Start’s effect by comparing the Head Start children with another group of children which has been indirectly improved by the Head Start children themselves.

This is certainly a possibility. However, most of the previous before-after studies of Head Start’s cognitive effects have shown, at most, small gains—so small that it is hard to imagine their having such major secondary effect on teachers and peers. Moreover, the first-grade children in the Westinghouse study were tested during the early part of their first-grade year—prior to the time when such secondary influence on teachers or peer children would have had a chance to occur. In results of direct measurements of the children (Metropolitan Readiness Test, Illinois Test of Psycholinguistic Abilities, and the like), there were only marginal differences between the Head Start and control-group children at that time. Also, on the Children’s Behavior Inventory, an instrument which obtained teachers’ ratings of the children, there were few significant differences between the two groups, indicating that the teachers were not able to perceive any differences between the motivation of the Head Start and non-Head Start children. In the light of these findings, it is hard to see how spillover or secondary effects could have occurred to an extent which contaminated the control group.
AN ASSESSMENT

Our over-all assessment of the study is as follows:

(1) In terms of its methodological and conceptual base, the study is a relatively good one. This in no way denies that many of the criticisms made of the study are valid. However, for the most part, they are the kind of criticisms that can be made of most pieces of social science research conducted outside the laboratory, in a real-world setting, with all of the logistical and measurement problems that such studies entail. And these methodological flaws open the door to the more political issues. Thus, one needs not only to examine the methodological substance of the criticisms which have been made of the study, but also to understand the social concern which lies behind them as well. Head Start has elicited national sympathy and has had the support and involvement of the educational profession. It is understandable that so many should rush to the defense of such a popular and humane program. But how many of the concerns over the size of the sample, control-group equivalency, and the appropriateness of covariance analysis, for example, would have been registered if the study had found positive differences in favor of Head Start?

(2) The scope of the study was limited, and it therefore failed to provide the answers to many questions which would have been useful in determining what specific changes should be made in the programs.

(3) Longitudinal studies, based on larger samples and covering a broader range of objectives, would be better, and should be undertaken. But until they are instituted, this study provides a useful piece of information that we can fit into a pattern of other reasonable evidence to improve our basis for decision-making. Thus, the Westinghouse study extends our knowledge, but does not fly in the face of past evidence. For the summer program, the study of a national sample shows what smaller studies have indicated—no lasting gain for the Head Start children relative to their peers. This may deflate some myths, but does not affect any hard facts. For the full-year program, the evidence of some limited effect is about as favorable as any we have found to date.

We imagine that this type of positive, but qualified assessment will fit any relatively good evaluation for some time to come. For we have never seen a field evaluation of a social-action program that could not be faulted legitimately by good methodologists, and we may never see one. But, if we are willing to accept real-world imperfections, and to use evaluative analysis with prudence, then such analysis can provide a far better basis for decision-making than we have had in the past.

What, then, does the Westinghouse study provide that will help in making decisions? First, the negative findings indicate that the program is failing, on the average, to produce discernible school success for its participants. Put more bluntly, the study shows that along the key cognitive and affective dimension, the program is not working at all well. And from this, one can infer, directly, that we should search hard for, and test, new techniques to make learning gains in the Head Start classroom more permanent and, indirectly, that the years before and after Head Start should also be looked at carefully. Second, the evidence suggests the superiority of the full-year over the summer programs. Most of all, we believe that the value of the study consists in the credible, validating evidence which it provides that the honeymoon of the last few years really ought to be over, and that the hard work of finding effective techniques should start in earnest.

Thus, the study pushes policy-makers toward certain decisions, particularly those involving within-program tradeoffs—more experimentation and more full-year projects in place of summer projects. Yet, and this would be true no matter how good the study was, the evidence is not a sufficient condition for major program-decisions. The last statement holds even for the within-program choices (tradeoffs, but not over-all cutbacks) and takes on greater cogency when one seeks implications for decisions concerning the need for more, or fewer, resources. The evaluative evidence must be considered in the light of other pieces of information and various highly important political judgments. For example: How deleterious would a program cutback be for program morale, or for our commitment to increase the outlays going to the disadvantaged for education? Surely, no reasonable person would claim that evaluative evidence alone is sufficient. Rather, such choices ought to be political, in the broad sense of that term, with credible evaluative data—a commodity heretofore in short supply—being considered as one of the inputs in the choice process.