DISCUSSION

Richard J. Light, Harvard University

The paper delivered by Professors Heber and Garber raises a number of fascinating questions about how to interpret the evaluation of a social intervention. I would like to devote my discussion to several of these points.

First, the investigators should be strongly praised for conducting this intervention as a true experiment, using randomization of subjects between treatment and control groups. Nearly all work in preschool intervention has not been experimental in the statistical sense; rather, programs have been evaluated by post hoc studies. Such studies usually have to create an after the fact artificial control group consisting of children with similar characteristics to those in the experimental group. Thus, because randomization was not employed, one rarely has any serious degree of confidence in the results of these studies. The Heber and Garber work obviously involved a serious effort to carefully define the study population of interest, and then to randomize children into two groups.

My major criticism of this work centers around how the treatment was defined. On page 6 of their paper, the authors report that, "individualized prescriptive teaching techniques were employed." On one hand, that is a constructive approach, since no doubt the families in the study displayed a range of individual differences in the kinds of help they would benefit from most. But on the other hand, the lack of a precisely defined treatment limits very severely the generalizability of any positive results. To be specific, suppose the city of Cambridge, Massachusetts decided that since the results of the Milwaukee investigation were so promising they wished to institute the treatment. What should they institute? I am afraid that the answer is not forthcoming from the research report of the Milwaukee Project. To say that help should be offered to mothers with low IQs on an individual basis, depending upon their needs, does not clearly specify any treatment. Thus the external validity, or generalizability, of results is severely limited.

A similar difficulty is created for the investigator's ability to make inferences internal to the study. The results reported by Heber and Garber can be summarized as being essentially a smashing success. Intervening with low IQ mothers and their children seemed to lead to enormous IQ gains on the part of the children, relative to a randomized control group. But what precisely was done with these mothers and children? According to the presentation, different families had different forms of intervention, with stress on different kinds of facilities. Thus, if we had to attribute the enormous program success to a particular treatment feature, we might easily conclude the critical feature to be the sensitivity and excellence of the teachers and social service people employed in this project. It is a pleasure to congratulate these people on their excellent work. But, once again, the difficulty for other investigators is that a treatment defined as being "unique for every family" is not a treatment easily generalizable, except perhaps in the extremely limited case where the identical social service team is involved.

This question of treatment specifiability is my primary criticism, and should be viewed in the light of my earlier comments about how delighted I am to see a randomized study of this nature. Let me now move ahead to a series of short questions about other issues raised by a preschool intervention study of this kind.

The primary dependent variable in this study was IQ. The treatment group had IQs after several years of intervention that far exceeded those of the control group. But the value of IQ scores is, ultimately, their ability to predict reasonably well a child's later performance in school. It will be interesting to see whether the higher IQ scores achieved by the treatment group children are good predictors of their school performance. In other words, there is some chance that these elevated scores, while correct in that they were obtained honestly, have a lower degree of predictivity than do IQ scores of children who have not been in a treatment group. A very likely possibility is that the treatment children will do better in school than the control children, but nowhere near as much as their remarkably increased IQ scores suggest they might.

The paper presents mean IQ scores for the two groups over a period of five years. But no data on variance is given. I am sure that Professors Heber and Garber have these data, and I raise this question not as a criticism, but rather out of curiosity as to whether some subject by treatment interactions may exist in these data. That is, might it be possible that "the program" is superb for some children of low IQ mothers, but not for others? Is it further possible that we might be able to identify what kinds of children, with what kinds of features, benefit most from this Heber and Garber program, so that it might be optimally targeted to children who are most likely to benefit? Such a result would not be surprising or unique, as much research with Head Start, the preschool intervention program, has shown that different kinds of curricula differentially benefit different kinds of children.

I would like at this point to mention one substantive finding of a colleague at Harvard, Professor Burton White. In working with middle income children, whose mothers have IQs in the "normal" range, he has reported that the critical period for children developing skills that tie in with later school competence is the age range 10 to 18 months. Professors Heber and Garber's data indicate that the critical period when the control group children begin to separate from the treatment group children first occurs at about age 18 months. I wonder if this difference is due to some feature of the treatment being reported here, or rather might have something to do with the social class of the families being studied. I have no idea as to the answer but would be interested if the investigators felt this difference was due to social class differences in the families.

A last observation has to do with the feasibility of adopting the form of family intervention reported by Professors Heber and Garber on a widespread national scale. The exciting feature of their work is that if their results hold up in replicated studies elsewhere, they will have succeeded in demonstrating that it is possible to substantially increase the IQs of children by an intensive intervention program begun at a young age. The worrisome feature is the practical one of cost. While no precise dollar figures are reported in their study, it is clear that the cost of the intervention runs to several thousand dollars per child per year. It is necessary then to ask the question, how does this intervention stack up on a cost benefit comparison with other forms of preschool remediation, such as, for example, the television program Sesame Street? I believe that the cost of Sesame Street is less than one dollar per year per child. Without arguing that televised instruction confers greater relative benefits than the intensive intervention program discussed here, I believe it is necessary for policy makers to raise this question.

To tie up these comments, I believe that the report by Professors Heber and Garber has two clear strengths. First, it illustrates that intensive remediation offers hope for substantially improving the IQs of children from low IQ family backgrounds, and this is most promising. Second, by using a randomized approach, it gives us greater confidence that the positive findings are not due to some artifact such as self selection, which would ultimately negate the value of these optimistic findings.

On the less optimistic side, the lack of a clear specification of the treatment, and the fact that the treatment varied quite a bit from family to family, makes it very difficult to generalize the results of this study. Science depends upon a steadily accumulating body of evidence from which, over time, we can draw stronger and stronger inferences. But it is very questionable whether a treatment that is not precisely defined can be replicated in different times and in different places, and this difficulty holds up scientific progress. Finally, I would urge that Professors Heber and Garber, together with other investigators who may decide to extend these findings, examine particularly carefully the question of subject by treatment interaction. A growing body of evidence in preschool education suggests that different types of curricula are best for different kinds of children. Discovering these interactions enables us to direct particular programs to the children who are most likely to benefit, and thus helps us organize effective social policy.