SCHOOL of PUBLIC POLICY

20
The Milwaukee Project

Douglas J. Besharov
Peter Germanis
Caeli A. Higney
and
Douglas M. Call

September 2011

Maryland School of Public Policy
Welfare Reform Academy
www.welfareacademy.org

Part of a forthcoming volume
Assessments of Twenty-Six Early Childhood Evaluations
by Douglas J. Besharov, Peter Germans, Caeli A. Higney, and Douglas M. Call

Note: This report is open to public comments, subject to review by the forum moderator. To leave a comment, please send an email to welfareacademy@umd.edu or fill out the comment form at http://www.welfareacademy.org/pubs/early_education/chapter20.html.

20
The Milwaukee Project

The Milwaukee Project operated in Milwaukee, Wisconsin, from 1966 to 1972. It was premised on the belief that the major factor behind the observed decline in IQ for low-income children as they grew older was not poverty, but rather the failure of parents to provide a stimulating environment for their children.[1] Those children whose parents had low IQs themselves were thought to be particularly at risk. The purpose of the project was to offset the progressive decline in IQ of infants identified as at risk for such declines. In order to do so, the project provided an intensive, six-year program offering services for both mother and child, including a year-round center-based educational program for children, beginning as soon after birth as feasible (usually within six months).

Howard Garber and his colleagues at the University of Wisconsin-Madison (the "UWM team") evaluated the Milwaukee Project using a random assignment-like process. The evaluation found large IQ effects at age five and at the end of eighth grade, but few lasting impacts on achievement. One possible explanation for the IQ effects, particularly during the intervention phase of the program, is that the children were taught skills that were similar to the items found on IQ tests for young children. Moreover, the small sample and high level of attrition raise many questions about the credibility of the findings.

Program Design

Program group. The Milwaukee Project targeted children in low-income, inner-city families whose mothers had IQs below 75. Indeed, children from low socioeconomic status (SES) families are more likely to have low IQs. The UWM team, however, compared the IQs of children in a low-income area of Milwaukee with the IQs of their mothers and fathers and found that this risk was especially acute in those families in which one or both parents had low IQs. They reported that nearly four-fifths of the children with IQs of 80 or under had mothers with IQs of 80 or under. It was believed that such parents deprived their children of an intellectually stimulating environment in the home. As Garber explains: "Mothers not only provide a genetic contribution

---

[1]Garber observed that the SES variable may be an imperfect proxy for the risk of retardation: ". . . the summarizing nature of SES undoubtedly conceals a considerable range of variation in maternal language skills and teaching styles and care should be taken not to equate disadvantagement with poor family functioning." *See* Howard L. Garber, *The Milwaukee Project: Preventing Mental Retardation in Children At Risk* (Washington, DC: American Association of Mental Retardation, 1988), 19.

for low intelligence, but may also influence their children's intellectual development by inadequately mediating early experiences for them."[2]

Most of the parents participating in the project were born in the South and had lived in Milwaukee for about five years, with an average of between nine and ten years of schooling. The average age of the mothers at the time of the participating child's birth was twenty-six years. The mothers had been raised in families with an average of 9.3 children and had an average IQ of 68. The fathers were about five years older and tended to work in manual jobs. These families were described as "substantially disadvantaged."[3]

Services. The Milwaukee Project was an intensive, six-year program that provided services to both mother and child. It included an extensive, daily, year-round education program for children, beginning as soon after birth as feasible (usually within six months) and continuing until the child entered school (around age six). Children attended an Infant Stimulation Center seven hours per day, five days per week. The program was intended "to provide the infants with the same quality of adult-child interaction that is usually found in the homes of normally functioning parents."[4] The center was staffed by paraprofessionals who had no formal training in child development, but were selected on the basis of being "warm, caring, and language facile adults."[5] During infancy, the adult-child ratio was 1:1 and was gradually increased to 1:3 by age two. As the children grew older, they were given cognitively appropriate learning activities, taught by certified preschool teachers. For the mothers, the intervention involved vocational and social education, including job training, remedial education, parent counseling, and family crisis intervention.

The Evaluation. The evaluation was intended to be based on random assignment, but there were difficulties implementing it (described later). A sample of black, English-speaking mothers with infants between three and six months of age in high-poverty neighborhoods was screened for participation. Those with IQs of 75 or lower and whose infants showed no signs of "central nervous system pathology" during the first few months after birth were asked to volunteer.

Major Findings

[2]Garber, 1988, 19.

[3]Garber, 1988, 33.

[4]Howard L. Garber, James D. Hodge, John Rynders, Richard Dever, and Raja Velu, "The Milwaukee Project: Setting the Record Straight," *American Journal of Mental Retardation* 95, (5) (1991): 510.

[5]Garber et al., 1991, 510.

The evaluation found large IQ effects at age five through the end of eighth grade, but few lasting impacts on achievement. One possible explanation for the IQ effects, particularly during the intervention phase of the program, is that the children were taught skills that were similar to the items found on IQ tests for young children.

Cognitive. The evaluation examined impacts on IQ and achievement.[6]

IQ. The Milwaukee Project achieved extremely large and consistent IQ gains. By age five, the program group had a mean IQ of 120, 26 points higher than the mean IQ of the control group (see table 1). Although the magnitude of the gain diminished over time, it remained 10 points when the children reached eighth grade (IQs averaging 101 compared to 91).

Table 1. Milwaukee Project: IQ Effects

	IQ scores		
Age (years)	Program group	Control group	Difference (percentage point)
2	125	96	29
3	126	94	32
4	126	96	30
5	118	93	25
6	119	87	32
7	103	81	22
8	103	83	20
9	103	84	19
10	104	86	18
11	99	89	10
14	101	91	10

Source: Howard L. Garber, James D. Hodge, John Rynders, Richard Dever, and Raja Velu, "The Milwaukee Project: Setting the Record Straight," *American Journal of Mental Retardation* 95, (5) (1991): 515.
Notes: Only significant differences are reported. IQ scores between the ages of two and six years were measured using the Stanford-Binet test; the Wechsler Intelligence Scale for Children (WISC) was used between the ages of seven and fourteen years.

[6]Garber et al, 1991, 510.

Achievement. Academic achievement was assessed after the first four years of school using the Metropolitan Achievement Test, but there was no difference in mean scores between the groups.[7]

School readiness/performance. By fourth grade, the program group had less grade retention (29 percent compared to 56 percent) and less placement in special education (41 percent compared to 89 percent). Despite the magnitude of these differences, however, they were not statistically significant.[8]

Socioemotional development. Relevant tests apparently not administered or results not reported.

Health. Data apparently either not collected or not reported.

Behavior. Data apparently either not collected or not reported.

Crime/delinquency. Data apparently either not collected or not reported.

Early/nonmarital births. Data apparently either not collected or not reported.

Economic outcomes. Data apparently either not collected or not reported.

Effects on parents. Although the intervention was designed to help mothers become employed by offering on-the-job training and adult education classes, "The reading levels for the mothers did not improve significantly, and they were only a little better able to acquire and hold jobs than were control mothers."[9]

Benefit-cost findings. Apparently a benefit-cost analysis was not performed. The UWM team apparently does not mention the cost of the program.

Overall Assessment

The Milwaukee Project was evaluated by the UWM team using a random assignment-like process. Moreover, the small sample and high level of attrition raise many questions about the

[7]W. Steven Barnett, "Long-Term Effects of Early Childhood Programs on Cognitive and School Outcomes," in *The Future of Children: Long-Term Outcomes of Early Childhood Programs* 5, no. 3 (Winter 1995): 31, http://www.futureofchildren.org/usr_doc/vol5no3ART2.pdf (accessed September 19, 2002).

[8]Barnett, 1995, 31.

[9]Garber et al., 1991, 509.

credibility of the findings.

Program theory. The UWM team notes, "The Milwaukee Project was based on the theory that declines in the rate of intellectual development were familial in nature rather than being a social phenomenon."[10] Thus, the Milwaukee Project was premised on the belief that the major factor behind the observed decline in IQ for low-income children as they grew older was not poverty, but rather the failure of parents to provide a stimulating environment for their children.[11] Those children whose parents had low IQs themselves were thought to be particularly at risk. The project sought to offset the progressive decline in IQ of infants identified as at risk for such declines. Within this context, the evaluation is appropriate.

Program implementation. The program appears to have been implemented without serious problems.

Assessing the randomization. A total of eighty-two mothers with IQs at or below 75 were identified, but only fifty-five met the criteria for participation. Although some eligible families moved or refused to participate, the main reason for the drop-off was that the infants had passed the six-month cutoff point for enrollment in the project. It is unclear how these changes affected the representativeness of the sample.

The project was initially contemplated as a random assignment evaluation, but practical considerations made it difficult to randomly assign families using a "traditional one-by-one randomization procedure." Instead, three or four families were selected each month or two and assigned to program and control groups on an alternating basis.[12] The researchers reported that the baseline characteristics of the program and control groups were comparable, but the number of characteristics examined was relatively small, and the small sample size means that observed differences would have to be quite large to be statistically significant.

Assessing statistical controls in experimental and nonexperimental evaluations. The evaluation used a random assignment-like procedure, so selection bias should not be a serious problem. However, no attempt was made to adjust for attrition.

---

[10]Garber et al., 1991, 498.

[11]Garber observed that the SES variable may be an imperfect proxy for the risk of retardation: ". . . the summarizing nature of SES undoubtedly conceals a considerable range of variation in maternal language skills and teaching styles and care should be taken not to equate disadvantagement with poor family functioning." *See* Howard L. Garber, *The Milwaukee Project: Preventing Mental Retardation in Children At Risk* (Washington, DC: American Association of Mental Retardation, 1988), 19.

[12]The UWM team explains that difficulty in finding eligible families and delays in opening the facility created problems in retaining the trained teaching paraprofessionals, so they began by assigning more subjects to the program group. *See* Garber et al., 1991, 497.

Sample size. Although fifty-five children were identified for the project, the analysis sample consisted of just thirty-five children. With such a small sample, extremely large impacts were needed to produce statistically significant findings. Indeed, a number of large differences between the program and control groups were not statistically significant. For example, the reductions in grade retention (29 percent vs. 56 percent) and placement in special education (41 percent vs. 89 percent) were very large, but were not statistically significant. Thus, the absence of impacts in some areas does not mean that the program did not affect some outcomes, but that the impacts may have been too small to be detected with the sample size. A small sample also means that differences in baseline characteristics would also have had to be very large to be statistically significant, making it more difficult to assess the comparability of the program and control groups.

Attrition. The UWM team conducted statistical tests on the original fifty-five families and report that "the assignment of these families did not vary significantly from random assignment."[13] However, fifteen families were lost after assignment, but before (or very soon after) participation, and five more were lost during participation. Garber concludes, "No significant differences between the experimental and control families were identified among these screening and assignment criteria nor did it seem that attrition significantly affected group means for these variables."[14] Again, given the small sample size, the differences would have to be very large to reach statistical significance.

Several issues raise questions about this conclusion. Although the UWM team presented information on the maternal IQs for the fifty-five families that were originally assigned, they only present detailed information on the family and infant characteristics of the forty families that actually enrolled in the project (and the thirty-five who completed the program). Thus, a full assessment of attrition was not conducted. In addition, the UWM team notes that there was "considerable missing data" and that they had to "substitute for missing values."[15] This substitution was based on group means. This procedure may have introduced biases as well.

Data collection. The data collection relied on a various standardized tests and parent surveys. The data sources are appropriate for the questions being studied, but are limited to a narrow set of outcomes.

Measurement issues. The evaluation relies on standard cognitive and achievement tests.

Generalizability. Only 5 percent of adult women in the general population have IQs of 75 or less, so the findings from the project, absent any concerns about the conduct of the evaluation,

[13]Garber et al., 1991, 506.

[14]Garber, 1988, 33.

[15]Garber et al., 1991, 512.

are limited to a narrow group.

Replication. The Milwaukee Project has not been replicated, and only the Abecedarian Project (see chapter 1) approached its intensity of services. They are the only two projects to find long-term IQ gains, although the Milwaukee Project's other findings, such as the effects on achievement tests, were disappointing.

Evaluator's description of findings. The UWM team explains that the surprisingly large IQ findings resulted primarily from the prevention of a decline that would otherwise have occurred in the absence of the intervention:

> Although the rate of development for experimental children remained relatively stable after 22 months of age, the decline experienced by control children continued up to about 30 to 36 months of age, then leveled off. The difference in the rate of development for these two groups increased to 29.20 IQ points by 30 months of age, after which it remained relatively stable for the remainder of the intervention phase of the project. This means that all of the divergence between the groups occurred during the infancy and toddler phases, the period when the intervention was least structured.[16]

Some observers are concerned that activities included in this (and similar) interventions teach developmental skills that are similar to those found on IQ tests. For example, in a 1989 review of the project, Arthur Jensen, an educational psychologist at the University of California at Berkeley, notes the lack of impacts in reading achievement test scores at fourth grade as an indication of this:

> The E-C [experimental-control] difference in IQ was not reflected in the nonsignificant E-C difference in Reading achievement scores or the questionably significant difference in Math achievement scores, on which by the end of the fourth grade, the mean scores for both the E and C groups were at about the 10th percentile of the normative sample. These results are most plausibly interpreted as a training effect of the intervention on the item content of the IQ tests without producing corresponding change in *g*, the general intelligence factor common to all cognitive tests, that the IQ ordinarily reflects in the untreated population.[17]

The UWM team seems to accept the validity of these concerns:

> We do not disagree that some of the infant activities were similar to the items found in the

[16]Garber et al., 1991, 513.

[17]Arthur R. Jensen, "Raising IQ Without Increasing G? A Review of The Milwaukee Project: Preventing Mental Retardation in Children at Risk," *Developmental Review* 9 (1989): 234–258.

> scales used to measure early intellectual development, but those of us who were in charge of developing the curriculum deny that there was any intent to directly teach the tests. . . . it would be impossible to develop a stimulation program for infants without including activities measured by these instruments. It is relatively easy to find specific content within this stimulation program that was similar to the item content of standardized tests. However, the real question is not "Did the intervention teach the test?" but "Did the intervention teach something that is not normally taught naturally every day by their own parents?"[18]

Moreover, while this explanation pertains to differences observed during the preschool years, it is less likely to explain differences much later. Unlike most other preschool projects, the IQ gains in the Milwaukee Project persisted into adolescence, averaging 10 points at age fourteen (when the last follow-up was conducted). There were, however, no statistically significant impacts on achievement tests.

The UWM team explains this anomaly by noting that intelligence is only partially correlated with achievement (at the 0.65 level),

> which means there is considerable variation in achievement that cannot be accounted for by intelligence. Children with low IQs almost always do poorly in school, but children with high IQs may cover the entire range of achievement from poor to excellent, depending on a number of other influences.[19]

They argue that intelligence is a "necessary but nonsufficient condition for achievement."[20] They further contended that this failure to produce achievement gains did not mean the project was ineffective with respect to its main objective—preventing intellectual retardation.

Garber argues that the preschool children's schools "failed to translate the academic potential of [preschool] children into the higher achievement levels to which the children seemed capable."[21] Since project staff helped move some of the preschool children into better schools, it is even more remarkable that the achievement gains did not persist as the children grew older.

Evaluator's independence. Howard Garber was the director of the Milwaukee program.

---

[18]Garber et al., 1991, 511.

[19]Garber et al., 1991, 519.

[20]Garber et al., 1991, 520.

[21]Garber, 1988, 277.

Statistical significance/confidence intervals. Statistical significance was measured and reported at the 1 and 5 percent levels.

Effect sizes. Apparently effect sizes were either not calculated or not reported.

Sustained effects. The evaluation examined impacts through age fourteen, nearly ten years after program completion.

Benefit-cost analysis. Apparently not performed.

Cost-effectiveness analysis. Apparently not performed.

Commentary

Editor's Note: For each evaluation included in this report, we attempted to contact the senior evaluators to offer them the opportunity to respond to our assessment. In the case of the Milwaukee Project, we contacted Dr. Howard Garber; however, he never responded to our request for comments.

Note: This report is open to public comments, subject to review by the forum moderator. To leave a comment, please send an email to welfareacademy@umd.edu or fill out the comment form at http://www.welfareacademy.org/pubs/early_education/chapter20.html.