1965. 41 years ago. I suspect many of you were not ready to submit Head Start research grant proposals when the program began! In fact, it’s hard to remember what the world was like 41 years ago, even if you were alive then. I was still a year away from completing my Ph.D. and was narrowly focused on testing Hull-Spence learning theory in an experiment with young children in Cedar Rapids, Iowa. Most of you probably know only second hand about events of the mid-1960s: the “British Invasion” of 1964-1966 that found the Dave Clark 5, the Beatles, and the Rolling Stones on our shores and airwaves. Bob Dylan penned “Like a Rolling Stone” at the beginning of the summer that Head Start began serving 561,000 children in 2,400 communities, and Sergeant Shriver, the first director of the Office of Economic Opportunity (OEO), believed he had found his “direction home.”

It’s hard to remember what an amazingly eventful year 1965 was:

---

1 In this version of my closing keynote presentation, I have added reference citations and some footnotes, which provide additional commentary or lead to historical resources for those who may be interested. I have incorporated the slides into the text where they are important for the presentation. Email contact for questions and/or comments: jlove@mathematica-mpr.com
• On January 4, President Lyndon Johnson proclaimed the Great Society in his State of the Union address.

• In February, the U.S. began bombing North Vietnam in operation Rolling Thunder.

• Feb. 21: Malcolm X was assassinated.

• March 7 became known as “Bloody Sunday,” when civil rights marchers began their trek from Selma to Montgomery and were violently confronted by state troopers.

• A month later, in a one-room schoolhouse in Texas, President Johnson signed into law the Elementary and Secondary Education Act of 1965, which we all know in its latest incarnation as No Child Left Behind.
• On August 11, the Watts riots began.

• That year, in Britain, Winston Churchill died—and J. K. Rowling was born. At the Grammies, *Mary Poppins* won best recording for children.

So with a spoonful of sugar and a ton of sweat and tears, a vast new program for children and families was underway.²

Within three years, a new head of the Office of Economic Opportunity (OEO) would have Head Start as part of his responsibility. His name: Donald Rumsfeld. He was responsible for guiding the country through an early phase of a different war—the War on Poverty.

---
² Photograph from an anonymous Head Start classroom in Buffalo, NY, circa late 1960s.
Research began immediately. One of the first studies in the new Head Start context that I could find is a dissertation by an enterprising young University of Chicago doctoral candidate by the name of Diana Slaughter, who investigated the maternal antecedents of African American Head Start children’s academic achievement. She has presented at this conference many times. One reason we don’t hear about that study and the other 506 that came out in the first six years of Head Start³ is because everybody—researchers, program administrators, and politicians—had to cope with the fallout of the so-called “Westinghouse report.”

In reflecting on my own involvement in Head Start research since the early 1970s, I’ve tried to think about lessons for the research we’re doing now and the research to come. I see moments of brilliance, and I see some ventures into blind alleys. I see important themes in the 40 years of Head Start research. After three days of conferencing, you don’t want to hear another research report or a summary of research findings, though I will mention some findings. But I’m going to focus on the Head Start research process. I will be selective, leaving out parts that others might have emphasized. I look forward to your comments at the end.⁴

A minute ago, I said that the summer Head Start program served 561,000—and that’s the magnitude of effort everyone remembers. And not without reason. Head Start must be considered the Ferrari of social programs, going from zero to 561 thousand in just a few months. But when September rolled around, the first year-long program enrolled … 20,000 children (Figure 1). And that so-called year-long program varied widely, from lasting just 3 to 4 months in some grantees to others that went 7 or 8 months (Smith and Bissell, 1970, p. 57).

---


⁴ For a recent, comprehensive review of Head Start research, see the chapter colleagues and I prepared for the new Blackwell Handbook of Early Childhood Development (Love, Tarullo, Raikes, and Chazan-Cohen 2006).
The majority of teachers had nothing more than a 6-day training program several colleges pulled together; one-third had no preschool teaching experience. Summer programs also served children who were older than those in the full-year programs: only 15 percent were under 5.

Research in the Early Years

So this is the program that was soon subjected to a major accountability test.\textsuperscript{5} In 1968-1969, the Westinghouse Learning Corporation and Ohio University located Head Start graduates in first, second, and third grades, matched them with children who had not been in Head Start, and drew their comparisons. One of the major legacies of the Westinghouse report is the perception that Head Start effects fade out by third grade. But this study was limited to cross-sectional comparisons. Third graders, who had been through the rocky summer and first year of Head Start, were compared with first graders who came along later. The poorer performance of the third graders was described as a “fade-out effect,” even though there were no data on the effect

\textsuperscript{5} Accountability was, indeed, the order of the day. According to Datta (1976), Congress and the Bureau of the Budget demanded accountability. The planning, programming and budgeting systems (PPBS) that Robert McNamara brought to the Pentagon influenced most federal agencies in the 1960s.
for these children before it faded. I wonder the extent to which the “myth of fade out,” as Steve Barnett has called it, stems initially from a study that was not designed for answering that question.

But many other studies took place in the early years. Their diversity foreshadowed the range of research that we see in the 21st century. Some examples are:

- University-based Evaluation and Research (E&R) Centers that conducted coordinated studies from 1967 to 1969. But because the interest was in within-program variation, no control groups were included.

Ray Collins, who was central to the Office of Child Development’s program planning in the 1970s and 80s, pointed out that the E&R Centers did not do rigorous impact evaluations. Rather, their strength was probing “inside the ‘black box’ of the total program in an attempt to gain insights into ways programs might be better designed” (Collins 1981, p. 52).

- The Educational Testing Service’s longitudinal study followed about 1,600 children in four sites from the time they were 3 ½ through third grade (1968-1973).

- In addition, there was a “profusion” (as Raizen and Bobrow called it) of individual, small-scale, often single-site studies—but they never were adequately synthesized.

Although Westinghouse was the biggie (politically), the most rigorous quasi-experimental study of Head Start’s early years was reported just last year. Jens Ludwig at Georgetown and Douglas Miller of UC Davis collaborated to apply an ingenious regression discontinuity analysis to data from Head Start’s first decade (Ludwig and Miller 2005).

Because OEO provided technical assistance to the 300 poorest counties in the country to help them prepare proposals for Head Start funding, Ludwig and Miller looked for

---

6 The Head Start Research Advisory Council, which Head Start Director Julius Richmond had instituted, did not disagree with OEO’s evaluation division about conducting summative evaluations, but it wanted a more balanced collection of studies. As Bronfenbrenner (1979) recalled, a research committee composed of Ed Zigler, Edmond Gordon, and Urie Bronfenbrenner was unsuccessful in arguing for an alternative to the proposed Westinghouse/Ohio evaluation because “it was the president’s decision that all departments of the Office of Economic Opportunity were to undertake cost-benefit analyses . . . of all their programs. Head Start could not be an exception” (Bronfenbrenner, 1979, p. 88).
“discontinuities in Head Start funding at the OEO cutoff for funding eligibility” (Ludwig and Miller 2005, p. 11). Using OEO’s cutoff, they found a large discontinuity for Head Start funding but not for other federal social spending. Applicants in the counties above the cutoff (the “treatment” group) were much more likely to receive Head Start funds, and this funding difference “persisted through the late 1970s” (p. 3). The authors argue that the results probably apply to Head Start programs between 1965 and the late 1970s, showing that:

- Mortality rates from causes that could be affected by Head Start—anemia, meningitis, and respiratory problems—were significantly lower as a result of Head Start funding.
- Head Start children were more likely to complete high school, by about 3 to 4 percentage points (based on Census data analysis).
- Former Head Start children completed about one-half year more of schooling than control children did (NELS2000 data).
- No effects were found on eighth grade reading or math scores (NELS 2000 data).

It’s no wonder that health-related effects were found. Look at what Head Start accomplished in its first two years, according to OEO:7

- 98,000 children with eye defects were discovered and treated.
- 900,000 dental cases were discovered, with an average of 5 cavities per child.
- 740,000 children were discovered without vaccinations against polio and were immunized.
- More than 1,000,000 were found not to be vaccinated against measles and also were immunized.

In a few minutes we’ll see why these numbers are important today.

---

7 Data from a March 1967 OEO “fact sheet,” as reported by Smith and Bissell (1970, p. 58).
Back in the 1960s and 70s, research continued at a rapid pace, in spite of—or perhaps because of—the criticisms of the Westinghouse study.\(^8\) As the responsibility for Head Start shifted from OEO (which had been in the Office of the President) to the Office of Child Development in the Department of Health, Education, and Welfare, program improvement and research were taken very seriously. OEO was spending between 1 and 2 percent of Head Start’s $340 million budget on research and evaluation (Vinovskis 2005, p. 103). That’s a lot. For perspective, if ACF spent what seems like such a small percentage on research today, we would see an annual budget of $100 million or more for Head Start-related research.

I’ve spent some time on ancient history because I worry about history repeating itself. We easily denounce the Westinghouse study—as though we couldn’t imagine anyone doing an *ex post facto* study these days. But when I see more and more studies using the ECLS-K data set to answer questions about Head Start and other preschool programs’ “effectiveness,” I get a little worried. To give just a few examples, in just the past year, we’ve seen such findings as:

- Children who had attended prekindergarten scored about a fifth of a standard deviation higher on a reading and math skills assessment at school entry than comparable children who spent the previous year at home (Magnuson, Meyers, Ruhn, and Waldfogel 2004).

- Entering preschool between ages 2 and 3 is better than either earlier or later for effects on later cognitive skills (Loeb, Bridges, Bassok, Fuller, and Rumberger 2005)

- Preschool programs lead to higher school achievement outcomes but more behavior problems in kindergarten (Rumberger and Tran 2006).

All this, when everything thing that’s known about the children’s preschool experience came from parents’ recollections at the beginning of kindergarten. Now, I’m sure these researchers

\(^8\) I highly recommend Lois-ellin Datta’s definitive analysis of the political-research context and aftermath of the study. Datta was Head Start’s national coordinator of evaluation from 1968 to 1972.
know more than I do about how to control for all the selection factors that may have affected which kinds of children attended which kinds of programs and at what ages and for how long, but I still worry. The Early Childhood Longitudinal Study has produced a wonderfully rich data set, and the studies I’ve cited explore important issues, but ECLS-K was not designed to answer causal questions about the effects of children’s experiences before kindergarten.

**Enduring Themes of Head Start Research**

I want to turn now to five themes in the history of Head Start research that I believe have currency today.

1. **The Shadow of Perry Preschool**

   Two years before Head Start began, 123 children participated in a preschool program that eventually made them the most policy-famous children on the planet. Those 123 children in Ypsilanti, Michigan, cast a large shadow over the 24 million who have gone through Head Start since then. There is no doubt that the cost-benefit results of the Perry study have had a powerful positive influence on public policies supporting early childhood education. Ironically, however, the very success of that experiment has had a negative influence in that it holds up a standard for expected effects that represent 1960s conditions and not 21st century community realities.

   For example, after one year of preschool the treatment-control impact on the PPVT had an effect size of .83 (Figure 2). The effect size was 1.22 for the Stanford Binet impact. As everyone knows by now, contemporary interventions like Head Start and Early Head Start have more modest-sized impacts, mostly under a quarter of a standard deviation. A number of writers seize on this contrast to say Head Start is not as effective as the older programs. I find that comparison troubling. The size of the impact of course depends on how well the program children did—but also on how well the control group children did. The context in the 1960s was vastly different
than it is today. Few alternatives to the intervention were available. Few states, for example, had universal kindergarten programs, let alone preschools. And the Perry children were a very disadvantaged sample. The control group entered the study with a mean PPVT score of 62\(^9\) and an average Stanford-Binet of 78.5. And remember the health statistics in those days, when three-quarters of a million kids in Head Start needed to be vaccinated against polio? And children had an average of five cavities?

Figure 2. Mean Peabody Picture Vocabulary Test Scores for Treatment and Control Groups\(^10\)

![Graph showing mean Peabody Picture Vocabulary Test scores for treatment and control groups.]

To compare effects across studies, we should at least try to approximate comparable samples. I’ll show what I mean with an example from the Early Head Start impacts on aggressive behavior problems (Figure 3). The effect sizes show the magnitude of the *reduction* in parent-rated aggressive behavior problems on the Achenbach scale. The larger the bar, the more that aggressive behavior was reduced. As you can see, the impact for African American

---

\(^9\) In contrast, the average Head Start entrant in fall 2000 scored 85 on the PPVT, according to a FACES report (ACF 2003); that is higher than the Perry Preschool intervention group’s mean at the end of preschool.

\(^10\) These data were reported in Weikart, Bond, and McNeil (1978); Figure 2 is adapted from their Figure 3 on p. 52.
children (the bar to the far right in the figure) was *three times* the size of the overall impact. If one wants to make comparisons with the Perry Preschool findings, we should look not at the .11 but at the .35, since Perry enrolled only African American children.

Figure 3. Which Impacts Should Be Compared Across Studies? Impacts (Effect Sizes in Percentages) for Reducing Aggressive Behavior Problems in the National Early Head Start Research and Evaluation Project\textsuperscript{11}

![Figure 3](chart.png)

I feel very privileged to have worked with Dave Weikart for eight years. What I’m trying to convey here is in no way a criticism of the Perry Preschool study but rather a sense of how it’s useful to view the “Perry phenomenon” today.

2. **The Median Is Not the Message\textsuperscript{12}**

With apologies to Marshall McLuhan, and while I have my Early Head Start slides handy, I want to make a point about subgroup analyses—how vital they are for understanding aspects of

\textsuperscript{11} Subgroup data are from ACF (2002): Table E.VI.11 (implementation of mixed-approach programs); Table VII.14, p. 391 (main language spoken at home); and Table VII.11, p. 382 (race/ethnicity).

\textsuperscript{12} Stephen Jay Gould originated this phrase as he poignantly fought the central tendencies doctors used to predict his cancer outcome (Gould 1985).
program effectiveness. Figure 4 shows the impact of Early Head Start on the percentage of parents reading to their 3-year-old every day. Again, the bars show the effect size of the impacts. The bar on the far left is the main finding we usually talk about—the small average impact for the total sample, with an effect size of 10 percent of a standard deviation.

![Figure 4. Example of Larger Effect Sizes (Percentages) in Subgroups: Percentage of Parents Reading Daily, National Early Head Start Research and Evaluation Project](image)

At the far right, we see the impact for the Early Head Start programs that implemented the Head Start performance standards early and took a mixed approach to serving families. Their impact on increasing the percentage of parents who read to their child every day was almost five times as large as the average impact. We can also see subgroups for whom the program was not very effective on this outcome (for example, the teenage moms shown in the shortest bar). We should not ignore the overall effect, but the subgroups tell more about the “what” and the “for**

---

13 Subgroup data are from ACF (2002): Table VII.8, p. 375 (mother’s age); Table VI.2, p. 275 (program approach); Table VII.17, p. 399 (risk levels); and Table E.VI.11, p. 311 (implementation of mixed-approach programs).

14 Another important subgroup finding, which I do not have time to discuss here, is the virtual absence of Early Head Start impacts on children with the highest degree of demographic risk. An excellent example of using subgroup findings to search for effective program strategies can be found in the work of Jane Knitzer and her colleagues at the National Center for Children in Poverty (Knitzer and Lefkowitz 2006).
whom.” These findings, for example, show policymakers the importance of performance standards, and suggest to programs that they need to find more effective ways of working with teenage mothers.15

3. The Measurement Dilemma: Are We Making Progress?

In February 1972, in my first year at High/Scope, Dennis Deloria and I flew from Detroit to Boston to spend a week at Abt Associates working with new colleagues on a proposal to conduct an evaluation of the Home Start Demonstration Program. I can still remember sitting at an IBM Selectric typewriter (in the days when “cut and paste” really meant something) struggling with the section on instruments to propose for measuring child outcomes. I wrote something to the effect that “while we can propose strong, highly reliable, and valid measures of children’s cognitive development, few such measures exist in the social-emotional domain.” I don’t know how many times I’ve written versions of that sentence since then! And we’re still discussing such issues at this conference.

Many have tackled this problem. In fact, in the late 1970s and early 1980s Head Start made a huge push to develop measures across all domains of development. The push actually began in 1974 when OCD commissioned the Rand Corporation to design a “national evaluation of the social competence effects of the Head Start program” (Raizen and Bobrow 1974, p. iii). One of the first things Rand did was to warn about “difficulties” in obtaining “interpretable and meaningful data” in any national evaluation without better child outcome measures. Rand was guided by the *OCD-Head Start Policy Manual*, which—thanks to the efforts of Ed Zigler, Clennie Murphy, Ray Collins, Jenni Klein, Sol Rosoff, and others—had just been published in

15 Complete results of the Early Head Start evaluation can be found in the project’s technical report (ACF 2002) and in a summary report recently published in *Developmental Psychology* (Love et al. 2005).
January 1973. This first systematic articulation of performance standards grew out of OCD’s Head Start Improvement and Innovation (I&I) effort. The performance standards clearly defined “social competence,” which became the centerpiece of Rand’s recommendations. It’s worth reading the definition, because in my experience the concept of social competence for Head Start has often been misunderstood—even since Ed Zigler coined the term 33 years ago:

Zigler defined social competence as:

“the child’s everyday effectiveness in dealing with his environment and later responsibilities in school and life. Social competence takes into account the interrelatedness of cognitive and intellectual development, physical and mental health, nutritional needs, and other factors that enable a child to function optimally” (quoted in Raizen and Bobrow 1974, p. 3).

The Head Start office had wanted the national evaluation to begin in fall 1974. Rand got a one-year delay for more planning, including test development (Raizen and Bobrow 1974, p. 4). But it was another 25 years before the national impact study got underway with the contract to Westat. But I’m getting ahead of my story.

After accepting the one-year delay, OCD funded a multi-year effort by Mediax Associates and several subcontractors to develop measures that would address the concerns Rand and others raised. Aptly called “the Head Start Measures Project,” it identified four domains for developing measures:

- Cognitive
- Social-emotional
- Health and physical development
- Applied strategies

This was more than 10 years before the nation’s governors created the first education goal with its five domains that were notable for including “approaches to learning” as a domain
parallel in importance to the cognitive, social, language, and health domains. As everyone in this room knows, when Head Start published its *Leaders Guide to Positive Child Outcomes* in September 2003, it took a similarly broad, comprehensive view, with a framework comprising eight broad domains (ACYF 2003). Clearly, Head Start has been leading the way.16

4. Planned Variation Is a Theme Whose Time Has Come—Again

In 2003, ACF’s Office of Planning, Research and Evaluation launched a new initiative called “Design Options for the Assessment of Head Start Quality Enhancements.” The basic concept is that when a new program idea arises, it is put to a rigorous experimental test before it’s widely disseminated. The Head Start Bureau (now the Office of Head Start) had been criticized for rolling out Early Head Start for tens of thousands of families before the programmatic approaches were tested. Although that criticism is only partially valid, a strong case can be made for getting new interventions “right” before requiring 49,000 teachers in 1,600 programs to change their practices. With a planned variation approach, one would try out, for example, a new plan for improving teacher training on a small scale, collect data using a rigorous experimental design, and then, if the results support it, spread the idea more widely.

The first planned variation study in Head Start, however, began 34 years earlier. In 1969, eight curriculum developers were funded to implement their curricula in multiple communities each. Results were, unfortunately, not very conclusive. This happened shortly after the U.S. Office of Education launched the Follow Through planned variation experiment. After the Head Start results were in, but before the Follow Through study was completed, Alice Rivlin and Michael Timpane convened a special meeting of the Brookings Panel on Social Experimentation

---

to assess what had been learned about doing such experiments. The title of the conference, and subsequent book, was *Planned Variation in Education: Should We Give Up or Try Harder?* (Rivlin and Timpane 1975). Recognizing the flaws in the two experiments, participants asked themselves, if we care about effective education for poor children, should we try to carry out better planned variation studies, or is the basic notion of planned variation doomed to failure? *Should we give up or try harder?* (Rivlin and Timpane 1975, p. 2, emphasis added).17

In the Head Start arena, it looks like ACF has decided to try harder.

In the 1970s, the Office of Child Development’s I&I initiative launched a series of demonstration programs that were sort of planned variation experiments. They studied variations but without the experiments. Still, new program ideas were tried out, typically in a limited number of program sites, through grants that augmented existing Head Start programs. Such demonstration programs were a major programmatic and research initiative of Head Start that had its heyday in the mid-1970s, but which have continued in every decade since. Some of them include:

<table>
<thead>
<tr>
<th>Program Type</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent and Child Development Centers (1970)</td>
<td></td>
</tr>
<tr>
<td>Health Start program (1971, “Healthy, That’s Me” curriculum)</td>
<td></td>
</tr>
<tr>
<td>Head Start Supplementary Training to improve personnel preparation</td>
<td>(1972)</td>
</tr>
<tr>
<td>Home Start Demonstration Program (1972)</td>
<td></td>
</tr>
<tr>
<td>Mainstreaming handicapped children pilot projects (1972)</td>
<td></td>
</tr>
<tr>
<td>Education for Parenthood (1972)</td>
<td></td>
</tr>
<tr>
<td>Child Development Associate credentialing program pilot projects (1973)</td>
<td></td>
</tr>
<tr>
<td>Child and Family Resource Program (1973)</td>
<td></td>
</tr>
<tr>
<td>Project Developmental Continuity (1974)</td>
<td></td>
</tr>
<tr>
<td>Basic Educational Skills demonstration (1978)</td>
<td></td>
</tr>
<tr>
<td>Comprehensive Child Development Centers (1988)</td>
<td></td>
</tr>
<tr>
<td>Head Start/Public School Transition Demonstration (1991)</td>
<td></td>
</tr>
<tr>
<td>Head Start Family Child Care Homes (1992)</td>
<td></td>
</tr>
</tbody>
</table>

---

17 I had not known until reading more about the Rivlin and Timpane conference for this paper, that the late David Kershaw, founder and first president of Mathematica, was a participant in that Brookings conference.
The Office of Child Development morphed into the Administration on Children, Youth and Families (ACYF), which continued trying out ideas in demonstration programs and evaluating them through the 1980s and 1990s. But these studies still did not address the basic questions program and policy folks had about the effectiveness of Head Start. That really began after the so-called “Blueprint Committee,” in 1990, urged Head Start to address questions about which Head Start practices maximize benefits for children and families with different characteristics under diverse circumstances (U.S. Department of Health and Human Services 1990).

5. Ambivalence About Randomized Trials

My perception is that Head Start has a love-hate relationship with experimental designs. I’ve worked on two experimental studies—Home Start in the ’70s and Early Head Start since 1995. But we didn’t see a national impact study of Head Start with a representative sample until the recent Head Start Impact Study. Of course, we have the FACES study, which came about in response to the government’s need for performance measures. For the better part of a decade, it has yielded very rich information about programs and their children in a nationally representative sample. This is very useful for knowing what happens in representative Head Start programs, but it doesn’t tell us about Head Start’s effectiveness.

We need diversity in research designs, and experiments are not always possible, but one of my concerns is that observational studies can produce vastly different findings from clinical trials—just think about the common medical wisdom regarding the benefits of hormone replacement therapies for post-menopausal women that was turned upside down by clinical trials demonstrating the risks of taking hormones.

Randomized trials are hard work, and we don’t always do them well. Just this spring, the results of a huge dietary study produced “disappointing” results in failing to show the benefits of low-fat diets for women. After eight years, the treatment and control groups had similar rates of
some cancers and cardiovascular disease. Then the clinicians observed that the intervention group did not reduce the fat content of their diets to the target levels—the treatment wasn’t really implemented. The investigators acknowledged the flaws in the execution of the study. We need to make the same acknowledgments when, for example, a home visiting program fails to show a difference in children’s school readiness—but we look at its implementation and see that the home visits occurred less frequently than called for and that many families left the program before their time was up. In the Head Start Impact Study, the intervention group didn’t take all the pills they were supposed to, and the control group got some of the same medicine.

These problems could just as well have occurred in a nonexperimental comparison design. But the rigorous protocol for a randomized trial increases the investigators’ vigilance for noncompliance. The main point I want to make is that those who do not think randomized designs are necessary should understand that design decisions can affect study conclusions and their policy implications.

6. Partnerships: This Conference’s Theme, and an Enduring Theme of Head Start Research

Partnership is the theme of this 2006 conference, but research-program partnerships have existed from the early days of Head Start research. At the time of the Home Start evaluation, we did not highlight the notion of partnership, but in 1976, Marrit Nauta and I wrote the following in the forward to the Home Start evaluation final report:

“Two people, especially, were responsible for the unusual potential that the evaluation design offered to those working on it….—the National Evaluation Project Officer, Dr. Esther Kresh, and the National Home Start [Program] Director, Dr. (Ruth) Ann O’Keefe…. Their close cooperation, beginning with the initial conceptualization of the National Home Start Demonstration Program in late 1971, led to a close-knit integration of evaluation and program activities and minimized most of the major problems typically faced by other evaluators on similar projects” (Love, Nauta, Coelen, Hewitt, and Ruopp 1976).
Head Start has spawned many research partnerships and consortia, including Head Start-University partnerships, the Child Outcomes Research and Support Consortium, the Early Promotion and Intervention Research Consortium, the Head Start Quality Research Center (HSQRC) Consortium, and the Interagency School Readiness Consortium. ACF research staff have collaborated with the National Institute of Child Health and Human Development (NICHD) on the Study of Early Child Care and Youth Development and the Early Head Start father studies and with the Department of Education on the ECLS-K and ECLS-B studies. And, of course, the Head Start’s eight research conferences represent a collaborative effort among many, many organizations and agencies.

In Early Head Start, we saw the most elaborate partnership I’ve ever experienced. ACF created a consortium that included program and research staff at the federal level, program staff and university researchers at the local level, the national evaluation contractor and local researchers (with program directors at times), and interactions with the national T&TA system.

The national and local researchers jointly created a publications policy that went beyond restrictions and encouraged collaboration. To this end, the consortium structure included nine working groups of like-minded researchers who have been collaborating on analysis and writing in areas as diverse as child care, fathers, risk and protective factors, school readiness, parenting processes, disabilities, and measures. These began forming in 1996 and are ongoing today, with some workgroups meeting tonight and tomorrow.

---

18 The latter is a joint effort of NICHD, ACF, and the Assistant Secretary for Planning and Evaluation (ASPE) in the U.S. Department of Health and Human Services and the Office of Special Education and Rehabilitative Services (OSERS) in the U.S. Department of Education. The initiative is designed to support research on the effectiveness of interventions, programs, and curricula in promoting school readiness for both typically and atypically developing children from birth through age 5.
One outcome of the research collaboration has been an incredible number of papers, presentations, book chapters, journal articles, and books. And a unique feature of the national impact reports was the integration of local researcher perspectives. The interim and final reports included appendixes devoted to brief local research reports—21 of them in a special volume of our final report, titled “Local Contributions to Understanding the Programs and Their Impacts.” In addition, if you read Volume I, the main report on our implementation and impact findings, you’ll find 22 full-page text boxes interspersed, authored by local researchers and summarizing local research studies that complement the findings being reported from the national evaluation.

I noted how the success of the Home Start Demonstration Program evaluation was at least partially a result of just a few people who were committed to collaborating. The same is true for the Early Head Start study. Without an associate commissioner for Head Start like the late Helen Taylor and an evaluation project monitor like Helen Raikes, the spirit of partnership would not have infused the collaboration activities we tried to carry out.

**Closing**

So we come to the end of these reflections—and the conference. We’ve seen a small part of the legacy from 41 years of Head Start research that continues to undergird today’s research. We have seen:

- Issues in comparing effects across studies, particularly from different eras
- Refining research questions and conducting subgroup analyses to learn how the program can be more effective, and to get away from simply asking *whether* it is effective
- The need—still—for measures that really get at the outcomes Head Start is committed to achieving
- The important role of experimental designs and use of planned variation studies
- The value added of research-research and research-program partnerships, and the importance of a few highly committed leaders
Head Start research has also confronted, dealt with, and sometimes helped to move us forward on a number of issues I have not touched on today, including:

- Concerns about assessing language-minority children
- Research aimed at understanding program implementation
- Program effectiveness for children with disabilities
- Research to understand how programs can be more effective with the highest risk families
- And many other areas

We now share a common heritage in this body of research that contains many lessons. But many challenges remain. Let me mention ones I think are particularly important.

1. **Learn More About Program Implementation**

   We have not solved the challenges of achieving fidelity of implementation when small-scale programs are taken to scale. In fact, there is evidence that we haven’t even solved the challenges of replicating a successful program on a small scale.

   Thirty years ago, Abt Associates completed a huge multi-curriculum, multisite, national, planned variation study of Follow Through without being able to conclude that one curriculum was clearly better than any other. The major problem was more variability across the sites where a particular curriculum was implemented than there was across the different curricula. On Monday, some of you attended the PCER poster symposium where one of the underlying themes related to the challenges of successfully implementing curriculum changes in public school preschools today. Clearly, we have a ways to go.
2. Use the Strongest Research Designs Possible

I don’t think our choice is simply to give up or try harder. We have to try smarter. In 1972, Lois-ellin Datta, who coordinated Head Start evaluations from 1968 to 1972, presented a review of the first six years of Head Start research at the NAEYC national conference. Her assessment of what’s lacking in our research base is blunt, and extreme. But I’m afraid it still rings true:

“Our country is fettered with programs that are dying from indifference: the data aren’t unfavorable enough to justify discarding them, aren’t clear enough to show how to modify them, or unequivocally favorable enough to justify expansion” (Datta, 1972, pp. 6-7).

I hope we’re in a more positive situation today. But evidence against that hope is seen in opposing reactions to the Head Start Impact Study. On the one hand, represented by Doug Besharov’s synopsis, we hear that:

“Confirming the findings of earlier, smaller evaluations, this new report found that Head Start has disappointingly small impacts on disadvantaged children” (Besharov 2005, p. 1).

On the other hand, represented by an SRCD policy brief,

“The positive effects of Head Start on children are comparable to or larger than those of other large-scale social programs” (SRCD 2005, p. 1).

Of course, both statements can be true, but the room for glass-half-empty-half-full interpretations suggests that the data on Head Start’s effectiveness are not yet clear and convincing enough.

3. Don’t Limit Head Start Research to Research on Head Start

This may seem like an odd suggestion, but even this conference follows this theme: It is Head Start’s research conference—presenting any research that’s relevant to and important for Head Start. As you know, Head Start pioneered research on the preschool-to-school transition
process with two major studies: Project Developmental Continuity in the 1970s (as part of the I&I program initiative I mentioned) and the Head Start Transition Study in the 1990s. One thing these studies accomplished was to make it clear that Head Start is not just about the lives of children when they are 4, or even 3 and 4. In fact, the Transition Study didn’t even collect data during the Head Start year. The issue is about preparing disadvantaged children for later schooling. We can’t learn all we need to know about Head Start and how it can be more effective if we don’t expand the scope of the research—and not merely by following the children after Head Start, but by learning more about their experiences before, during, and after Head Start.

4. Minimize the Role of Luck

Finally, we need to capitalize on research to minimize luck in determining children’s futures. That’s really what our research is about. The Arkansas poet Miller Williams (1997, pp. 69-70) read his poem, “Of History and Hope” at President Clinton’s second inaugural in 1997. I want to read an excerpt, and then close.

“But how do we fashion the future? Who can say how except in the minds of those who will call it Now?
The children. The children. And how does our garden grow?
With waving hands—oh, rarely in a row—
and flowering faces. And brambles, that we can no longer allow.
Who were many people coming together
cannot become one people falling apart.
Who dreamed for every child an even chance
cannot let luck alone turn doorknobs or not.”
In summer 1965, the Office of Economic Opportunity launched the first Head Start programs as part of the War on Poverty. Head Start has been the doorknob for 24 million children so far. Research in the early years included practically every design, measurement, and analytic challenge that program evaluators can imagine. The ways in which researchers dealt with those challenges provided lessons that have guided (and sometimes been ignored) in subsequent research and evaluation studies. By paying attention to those lessons, we can ensure more effective use of research and evaluation for program improvement and policy formation. The next 24 million children and families are depending on our research—our thoughtful and careful research—to increase their chances in life—by reducing chance.

Thank you.

References


