The Dubious Promise of Universal Preschool

David J. Armor and Sonia Sousa

In his 2013 state of the union address, President Obama proposed a “Preschool for All” initiative, pledging $75 billion in new federal funding over a period of ten years. In a partnership with the states, the federal government would provide the majority of funds needed for the implementation of “high-quality” preschool for all four-year-old children whose families make up to 200% of the poverty line. The federal government would also provide incentives for states to offer preschool to all remaining middle-class children, thus approaching “universal” preschool.

The idea of universal preschool is not new, and a handful of states already have such policies. Oklahoma, for instance, has the highest rate of four-year-old enrollment at about 70%. The president’s endorsement and the offer of federal start-up money give a tremendous boost to this concept. Before we launch a universal preschool program and dedicate billions of dollars in federal funds, however, it is important that we take into account the outcomes of students participating in the existing federal preschool program: Head Start.

The Head Start program has been evaluated using the most sophisticated research designs available to social scientists, and the results have been disappointing. While Head Start appears to produce modest positive effects during the preschool years, these effects do not last even into kindergarten, much less through the early elementary years. These findings

David J. Armor is professor emeritus of public policy at the School of Public Policy at George Mason University.

Sonia Sousa is an affiliate assistant professor at the School of Public Policy at George Mason University.
suggest that, if a new universal preschool program is to have greater success, something about the new program will have to be different.

Proponents of universal preschool stress that children need high-quality programs, implying that Head Start is not high quality and thereby explaining why the program has not produced the desired results. Evidence from comparative studies suggests, however, that Head Start programs are not significantly different from those commonly cited as being high quality, at least in attributes that produce better outcomes. The evidence suggests that we aren’t exactly sure what kind of program would succeed in leveling the playing field for low-income children. A review of the research on Head Start and other preschool programs, with a rigorous consideration of the benefits claimed by these programs and an informed cost-benefit analysis, must therefore be undertaken before the federal government spends billions of dollars on a new, untested program.

**The State of Head Start**

Head Start began in 1965 as a summer-school program intended to advance the ends of President Lyndon Johnson’s War on Poverty. Policymakers quickly realized, however, that an eight-week intervention was not enough to overcome the disadvantages children suffered after four years of poverty, so Head Start was converted into a full-time program that served fewer students. In the early years of the summer-school program, about 700,000 students enrolled at the relatively low cost of $2 billion, for a per-capita cost of $2,000 to $3,000 (all in 2011 dollars). After the conversion to a full-time program, enrollment dropped to under 400,000 by the early 1970s. The figure on the next page tracks the growth of program enrollment and costs from 1967 to 2011.

From the early 1970s to about 1989, Head Start enrollment was relatively stable, increasing from just under 400,000 to about 450,000 children. Funding was also relatively stable, growing from $2 billion to $2.3 billion, and per-capita spending hovered around $5,000 per child in current dollars. In 1990, the program began a decade of rapid growth, starting in the first Bush administration and continuing through the Clinton years. Enrollment more than doubled to just over 900,000 children, and funding climbed even more steeply; by 2000, total appropriations had tripled to just under $7 billion per year, and per-capita expenditures increased to $8,000 per child. These funding increases,
primarily meant to pay for curriculum changes and better-trained instructors, were partly in response to early evaluations of the Head Start program, which showed gains that were very modest and tended to fade after children entered elementary school.

**TRENDS IN HEAD START ENROLLMENT AND APPROPRIATIONS**

During the George W. Bush administration, enrollment remained stable around 900,000 children, but funding increased—jumping to $8.2 billion in 2002 and then slowly drifting downward to about $7.2 billion in 2008. In the first two years of the Obama administration, the American Recovery and Reinvestment Act added $2.1 billion to Head Start (divided between 2009 and 2010), and enrollment increased by 61,000. In 2011, program funding dropped back to $7.6 billion, serving approximately one million students.

The Head Start program was originally implemented to help children from low-income households catch up with middle-class children by the time they reached kindergarten. For decades, studies have shown that children who grow up in poverty are more likely to have social and behavioral problems, as well as cognitive challenges that can impede learning. Children who experience poverty in their preschool years, in
particular, have a substantially higher risk of not graduating from high school and so of suffering all of the resulting economic consequences. Head Start was intended to mitigate some of those disadvantages through early intervention. The program is based on a “whole child” model, aimed at improving four main contributors to a child’s readiness to enter regular school at age five: cognitive development; social-emotional development; medical, dental, and mental health; and parenting practices.

As part of a federal initiative under the Department of Health and Human Services, Head Start programs operate independently from local school districts and are most commonly administered through city or county social-services agencies. In general, classes are small or at least have low child-to-staff ratios of less than ten students per adult staff member. Individual Head Start programs develop their own curricula of academic and social activities with federal performance standards in mind. Though most Head Start teachers do not have a bachelor’s degree and are not certified teachers, most of the teachers do have at least an associate’s degree, and most have completed six or more courses in early-childhood education.

HHS demands that certain federal requirements be met in all Head Start programs, but because each program is run locally, there can be wide variations in the quality of the programs. In addition, Head Start does not require that teachers have the certification that regular public-school teachers are required to have.

As a result of these variations and trends, some champions of early-childhood intervention programs have argued that the Head Start program, in terms of staff and curriculum, is not a “high-quality” preschool program. A high-quality program, according to these critics, must have a fully certified teacher and a standard curriculum approved by a state or local school board. Head Start officials dispute this assertion and argue that its instructors and curricula both meet high standards.

**Does Head Start Kick-start Kids?**

After more than four decades of Head Start, there has been a considerable amount of research done on the program and its effects. The most comprehensive and rigorous evaluation to date is the Head Start Impact Study (HSIS), sponsored by the Administration for Children and Families in the Department of Health and Human Services. The study looked at
4,667 three- and four-year-old children applying for entry into Head Start in a nationally representative sample of programs across 23 states.

Children were randomly assigned to either a Head Start group (treatment group) or a non-Head Start group (control group), ensuring that the children in the two groups were similar in all measured and unmeasured characteristics at program entry, the only difference being their participation in Head Start. In order to verify this initial equivalence, and to estimate potential effects of attrition, pre-test assessments of all critical outcome measures were taken before the start of the preschool experience. Children in the non-Head Start control group could enroll in other non-Head Start child services chosen by their parents or remain at home in parent care. This allowed HSIS to assess how well the program performs compared to what children would experience in the absence of the program, rather than compared to an artificial condition where children were prevented from obtaining other child-care services. Nearly half of the control-group children did enroll in other preschool programs.

Outcome measures covered all four Head Start program goals: cognitive development, social-emotional development, health status and access to health care, and parenting practices. The table below summarizes only the positive effects on cognitive skills and social and behavioral outcomes (the health and parenting outcomes as well as some negative effects will be discussed separately).

### NUMBER OF SIGNIFICANT AND POSITIVE EFFECTS OF HEAD START ON COGNITIVE AND BEHAVIORAL OUTCOMES

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Head Start Year</th>
<th>Kindergarten</th>
<th>First Grade</th>
<th>Third Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive Skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Head Start at age 4</td>
<td>7 out of 14</td>
<td>0 out of 19</td>
<td>1 out of 23</td>
<td>1 out of 11</td>
</tr>
<tr>
<td>Head Start at age 3</td>
<td>8 out of 14</td>
<td>1 out of 19</td>
<td>1 out of 23</td>
<td>0 out of 11</td>
</tr>
<tr>
<td>Social and Behavioral Skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Head Start at age 4</td>
<td>0 out of 9</td>
<td>0 out of 20</td>
<td>1 out of 20</td>
<td>2 out of 20</td>
</tr>
<tr>
<td>Head Start at age 3</td>
<td>2 out of 9</td>
<td>2 out of 20</td>
<td>2 out of 20</td>
<td>1 out of 20</td>
</tr>
</tbody>
</table>


Relative to children in the control group, Head Start participants showed positive effects in numerous cognitive skills during their Head Start years including letter-naming, vocabulary, letter-word
identification, and applied math problems. However, these improvements did not carry into kindergarten or the later elementary grades. Only a few outcomes showed statistically significant improvements in the later grades, and they were not consistent across the three- and four-year-old cohorts (that is, they applied to different skills across cohorts and grades). This inconsistency makes it difficult to generalize about the success of the program in preparing students cognitively to keep up in later grades.

Participants showed fewer significant improvements in social and behavioral skills, even in the Head Start year, and results were inconsistent between the three- and four-year-old cohorts. The four-year-old cohort showed no significant improvements in the Head Start year or kindergarten, but in third grade they showed a significant reduction in total problem behavior according to their parents. The three-year-old cohort showed several significant improvements in social and behavioral skills, but only for outcomes assessed by parents. Not shown in the table, however, are several significant negative effects in relationships with teachers as rated by first-grade and third-grade teachers; in fact, there were no significant positive effects for this cohort as assessed by teachers for any of the elementary years. The three-year-old cohort showed several significant improvements in social and behavioral skills, but only as assessed by their parents.

Head Start also had some significant health-related effects, especially in increasing the number of children receiving dental care and having health-insurance coverage. These effects were not consistent, however. For example, participants did have increased health-insurance coverage, but not during their year in Head Start, and it did not extend into the third-grade year for either cohort, suggesting a possible short-term effect after leaving Head Start. There were also some significant effects on parenting practices, but they applied to the three-year-old cohort only. Most of these practices related to discipline, such as reduced spanking or use of time-outs. The reduced-spanking outcome occurred during the Head Start year and kindergarten, but it did not last into the first or third grades. During the Head Start year, there was also a significant effect on the amount of time parents spent reading to their children, but this effect did not last into kindergarten or first grade.

A secondary analysis of the HSIS evaluation was done by Peter Bernardy in his doctoral dissertation, entitled "Head Start: Assessing
Common Explanations for the Apparent Disappearance of Initial Positive Effects.” The study uses the HSIS data to test various hypotheses (including the quality argument) about why the initial effects of Head Start weaken and disappear during kindergarten and first grade. To begin, he examined whether other learning skills not examined in the HSIS might be affected more strongly than cognitive skills. These included child attention behavior, child persistence, and child confidence, as evaluated by teachers, parents, and independent assessors. Improvements in these skills could portend better longer-term outcomes in academic performance.

More important from a policy perspective, Bernardy also examined several programmatic and methodological issues frequently cited as explanations for why Head Start’s positive effects fade after students enter regular grade school. He tested whether the quality of the Head Start curriculum and the quality of the elementary schools attended by Head Start students after preschool — two of the most common excuses for the failure of Head Start programs — affected the outcomes of students. Also addressed was the methodological problem arising from the fact that a significant portion of the control-group students attended some sort of other preschool program, meaning there was no true “no preschool” control group.

The table on the next page summarizes the findings of Bernardy’s study. The results are grouped under the four issues addressed: possible additional learning skills, Head Start program quality, elementary-school quality, and a no-preschool control group.

For additional learning skills, there was only one statistically significant positive effect out of the 43 possible comparisons, and there were none in the elementary grades. This single significant effect was the parent rating of attention at the end of the Head Start year for three-year-old children. This outcome was offset, however, by non-significant findings for attention as rated by independent assessors and teachers.

Children who participated in Head Start did exhibit several significant positive effects compared to children who had no preschool at all. All of these effects, however, were limited to the year they spent in the Head Start program; there were no significant positive effects in the kindergarten or first-grade years. In fact, there were two significant negative effects shown in cognitive skills for the three-year-old cohort, meaning the control group that did not participate in any preschool program had higher cognitive scores.
**SIGNIFICANT AND POSITIVE EFFECTS OF HEAD START IN THE BERNARDY STUDY**

<table>
<thead>
<tr>
<th>Additional Learning Skills</th>
<th>Head Start at Age 3</th>
<th>Head Start at Age 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age 3</td>
<td>Age 4</td>
</tr>
<tr>
<td>Attention: teacher rating</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Attention: parent rating</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Attention: independent rating</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Attention: teacher/parent</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Persistence: indep. rating</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Confidence: indep. rating</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Confidence: teacher/parent</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

**Head Start Curriculum Quality**

| Cognitive skills | No   | No   | No   | No   | No(-) | No   | No   |
| Social & behavioral skills | No   | No   | No   | No   | No   | No   | No   |
| Attention: independent rating | No   | No   | No   | No   | No   | No   | No   |
| Persistence: indep. rating | No   | No   | No   | No   | No   | No   | No   |
| Confidence: indep. rating | No   | No(-) | No   | No   | No   | No   | No   |

**Elementary School Quality**

| Cognitive skills | --   | --   | No   | No   | --   | No   | No   |
| Social & behavioral skills | --   | --   | No   | No   | --   | No   | No   |
| Attention: independent rating | --   | --   | No   | No   | --   | No   | No   |
| Persistence: indep. rating | --   | --   | No   | No   | --   | No   | No   |
| Confidence: indep. rating | --   | --   | No   | No   | --   | No   | No   |

**Head Start vs. No Preschool**

| Cognitive skills | Yes  | No(-) | No(-) | No   | Yes  | No   | No   |
| Social & behavioral skills | No   | No   | No   | No   | No   | No   | No   |
| Attention: independent rating | No   | No   | No   | No   | No   | No   | No   |
| Persistence: indep. rating | Yes  | No   | No   | No   | Yes  | No   | No   |
| Confidence: indep. rating | Yes  | No   | No   | No   | Yes  | No   | No   |

Source: Authors' construction based on Bernardy (2012). Significant means p≤.05.
* A minus sign denotes a significant negative effect.
The failure of Head Start to produce long-term effects on its participants is often blamed on the quality of the Head Start curriculum. Using a common rating scale for assessing preschool curriculum quality, the Bernardy study showed no positive relationship between higher-quality Head Start programs and outcomes. Similarly, the evaluation of elementary-school quality yielded no significant positive effects after controlling for higher-quality schools.

This is why the Bernardy study is so valuable. It revealed no significant relationship between Head Start program quality and the major cognitive and social outcomes. And the long-term effects of participating in Head Start programs were not statistically different from not going to preschool at all.

**The Quality Question**

The evidence from the nationally representative HSIS evaluations, along with the Bernardy study, indicates that Head Start has small to moderate positive effects for several outcomes by the end of a child’s preschool experience, but these initial positive effects do not endure even through kindergarten, let alone through the third grade. Proponents of publicly funded preschool contrast Head Start with other preschool programs that have, allegedly, better long-term outcomes, blaming the difference on the lower quality of Head Start. Some of the better-known “high-quality” programs created and evaluated within the last ten years include the Abbot program in New Jersey, the Boston preschool program, a preschool program in Tulsa, Oklahoma, and a Tennessee preschool program. Evaluators of the first three programs — none of which used the rigorous randomized designs used in the HSIS — claim very large cognitive effects, some nearly ten times those of the typical Head Start program. These stronger effects are attributed to the quality of the program offerings, namely, a stronger curriculum and fully certified instructional staff with bachelor’s degrees. But the Tennessee preschool evaluation, which had a randomized design, failed to find significant positive effects through first grade.

The argument that the large effects noted in New Jersey, Boston, and Tulsa are due to better curriculum and more qualified instructional staff — particularly more teachers with bachelor’s degrees — is especially puzzling. Research has shown that both curriculum quality and teacher education have very low correlations with cognitive and
social-emotional outcomes in preschool programs. This was demon-
strated both in Bernardy’s study using the national sample of Head
Start programs and in a 2007 study by a team led by Diane Early of the
University of North Carolina, which used a diverse national sample of
preschool programs including Head Start. The lack of correlation makes
sense when one considers that the children in these preschool programs
are only three or four years old. Teaching basic vocabulary or numeracy
skills to this age group does not require years of formal study or a com-
plex curriculum, otherwise untrained middle-class parents would not
be such good teachers for their young children.

Furthermore, Head Start actually competes well on quality measures
with preschools that have reputations for high quality. According to the
ECERS-R index, a common measure of preschool classroom quality,
the Head Start programs studied in the HSIS are actually of slightly
higher quality than the “high-quality” Abbot program. The index
ranges from 1 (poor) to 7 (excellent), and the Abbot preschool averaged
4.8 in a 2007 evaluation study led by Ellen Frede. For the national Head
Start sample in the HSIS, the average ECERS-R index was 5.2 for the
three-year-old Head Start cohort and 5.3 for the four-year-old cohort.
(The Boston and Tulsa pre-K programs do not report ECERS scores.)

It is also important to note that there are potentially serious method-
ological problems in the New Jersey, Boston, and Tulsa evaluations.
Because the preschool programs deal with very young children, often in
challenging communities, it is difficult to achieve randomized designs.
The Abbot, Boston, and Tulsa studies, therefore, used a quasi-experimental
design called regression discontinuity design, or RDD.

There is no pre-test in RDD; the treatment group is assessed at the
beginning of kindergarten, after children have completed the preschool
program. The control group consists of students one year younger who
are just starting pre-K. The treatment group can experience attrition,
and children who drop out of a preschool program are quite likely to
have lower test scores. In effect, the treatment group consists of only
those students who complete a year of preschool, and we would expect
their scores to be higher (adjusting for age) than a control group that
includes everyone who starts the pre-K program. This could be a major
reason why RDD studies of preschools show much higher program ef-
facts than the HSIS. The results of the Tennessee program, in contrast,
look very similar to Head Start, with substantial effects in the Head Start
year but no longer-term effects. The Tennessee program had the same quality characteristics as the other “high-quality” programs; the only difference is the use of a rigorous randomized design with a pre-test.

Having rigorous studies of these preschool-program outcomes is important to ensuring that children are receiving the quality of education they need. Quality-comparison studies are vital to making sure that the low-income children that Head Start was designed to help are actually getting the boost they need to keep up with their middle-class counterparts in kindergarten. From a policy perspective, such studies are also important for cost-benefit analysis, especially as a new universal-preschool program is being considered.

**Costs and Benefits**

Only a handful of cost-benefit studies have been carried out for pre-K programs. There is no need for a cost-benefit analysis of Head Start given the available evidence, because the program has only short-term effects and no quantifiable long-term benefits. The newer high-quality programs discussed earlier are too new for a meaningful cost-benefit study, and their treatment groups may be biased by program dropouts, meaning that a valid evaluation of long-term costs and benefits may not be possible using these newer studies.

Two older “high-quality” preschool programs mentioned frequently in the research literature are the HighScope Perry Preschool program from Ypsilanti, Michigan, and the Chicago Child-Parent Center (CPC) program. These programs include parent education and support and thus differ in significant ways from the type of preschool programs offered by Head Start, as well as the more recent “high-quality” programs. They are also much more expensive on a per-capita basis. Both the CPC and the HighScope Perry Preschool programs were started in the 1960s, and researchers have carried out long-term cost-benefit analyses for both programs. These analyses conclude that better long-term outcomes more than pay for their higher program costs, mainly in the form of higher career income and lower rates of criminal behavior.

For several reasons, however, these programs may not be reliable indicators of the long-term benefits of current pre-K proposals. First, while the Perry Preschool results are based on a randomized design, the CPC study, led by Arthur Reynolds and Judy Temple, used a quasi-experimental group without a pre-test for major outcome variables.
In addition, the treatment group consisted of minority children who graduated from CPC preschool and kindergarten programs, while the control group is comprised of kindergarten students who participated in other pre-K and kindergarten programs. The treatment group could therefore be biased by program dropouts, just as in the newer programs’ studies with the RDD designs. Further, without a pre-test, there is no way to test or adjust for these potential differences between the treatment and control groups.

Second, while the Perry Preschool study is a state-of-the-art cost-benefit analysis conducted by a highly professional team of economists led by James Heckman, it is not without flaws when it comes to comparing the small 58-student treatment group and the 65-student control group. The randomized design was also compromised by re-assigning two parents from the treatment to the control group because they were single mothers who were working and could not participate in the program. In addition, four children in the treatment group dropped out of the program. The authors argue that the re-assignments were taken into account by matching techniques, but no mention is made of the program dropouts. As in the RDD studies discussed above, program dropouts may be the most at-risk children with the worst prognoses for long-term outcomes. While the number of dropouts is small, their omission from the long-term follow-up study could overstate treatment-group success.

Finally, the family-intervention components of these programs make it impossible to compare them to the high-quality programs envisaged by the Obama proposal. It also makes these programs very costly compared to Head Start and other regular pre-K programs that have been implemented by several cities and states. For example, in 2006 the per-pupil cost for the Perry Preschool program was nearly $18,000 ($20,000 in current dollars). It is unlikely that many states would agree to invest those types of funds in a program that was tested 50 years ago on a very small group of children from a single small city.

**Better Evaluation for More Cost-Effective Outcomes**

It would be similarly unwise for the federal government to invest $75 billion in universal preschool when it is not at all clear that Head Start or other preschool programs have positive long-term effects on participants. The HSIS evaluation, the Bernardy study of the same data, and the Tennessee study remain the most rigorous studies conducted to date.
on the effectiveness of preschool programs. These studies do not find preschool to be effective in increasing long-term cognitive or social and emotional outcomes. In cost-benefit terms, we cannot claim that the $8,000-per-child Head Start program is cost effective.

Despite the flaws in the evaluations of other “high-quality” programs, many education policy experts might nonetheless argue that some type of higher-quality pre-K program is needed for disadvantaged children. In response, we would urge President Obama to modify his proposal for universal pre-K. The president’s plan as it has been proposed would be modeled after the curriculums of existing pre-K programs that are part of regular school systems, and the new program’s teachers would be required to have certification. As we have seen, however, there is insufficient evidence to suggest that these changes will yield a program that will succeed where Head Start has failed.

Rather than implementing a full-blown program, the president should propose a national demonstration project for pre-K in a selected number of cities and states, accompanied by a rigorous randomized evaluation that would follow participants at least into the third grade. This demonstration project should also examine whether “preschool for all” closes achievement gaps between rich and poor, since it is possible that middle-class children will benefit more than disadvantaged children.

This randomized evaluation study should also assess what type of program characteristics are necessary for an effective yet affordable pre-K program. Such a randomized study should test the effect of each programmatic factor (staffing and curriculum characteristics, for instance) that other studies suggest is important in determining students’ cognitive and social-emotional outcomes. The cost implications are quite substantial, because the cost of pre-K as a regular program in a K-12 system may be considerably higher than the cost of Head Start.

Given today’s budget and deficit pressures, this national demonstration should be kept budget-neutral. Instead of new expenditures, a portion of current Head Start funds should be redirected to the national demonstration program and evaluation. Assuming a demonstration program for about 20,000 students spread across, say, 20 school districts, with average program costs of $10,000 per student, demonstration costs would be approximately $200 million per year; a rigorous evaluation should cost no more than one-tenth of that amount. With a current Head Start budget of about $8 billion, this demonstration project would
consume only 2% to 3% of Head Start funds. It might be possible to locate the demonstration in communities that lose Head Start programs so that there would be no major reduction in services for children currently using Head Start.

Such a demonstration project would be vital to enabling a truly high-quality universal preschool program to work, especially in a cost-effective way. There is simply no evidence to date that indicates a program like the one President Obama has proposed would produce meaningful, long-term positive outcomes for participants. There is also no evidence that such a program would be cost-effective. If the federal government is going to create a new program to provide broad access to high-quality preschool education, it should ensure that the program can produce positive long-term results without wasting valuable funds. American taxpayers and their children deserve at least that much.