

BACKGROUND

No. 3040 | AUGUST 06, 2015

The Head Start CARES Demonstration: Another Failed Federal Early Childhood Education Program

David B. Muhlhausen, PhD

Abstract

In search of evidence that Head Start can be an effective program, the U.S. Department of Health and Human Services initiated the Head Start CARES demonstration project. The Head Start CARES demonstration project tested three “evidence-based” social-emotional interventions to determine whether these interventions help disadvantaged children to develop appropriate social-emotional behaviors. The answer was no. Experimental evaluations released in 2014 and 2015 by the U.S. Department of Health and Human Services found that enhanced Head Start CARES demonstration programs had little to no effect, compared with regular Head Start services, on the social-emotional and academic skills of participating children.

Calls for the federal government to fund universal preschool programs and expand early childhood education programs seem to never cease. The two small-scale studies—of the High/Scope Perry Preschool Project begun in 1962 and the Carolina Abecedarian Project begun in 1972—that were used to demonstrate the effectiveness of such interventions are now outdated.¹ Their results have never been replicated.² No evidence indicates that these programs can produce the same benefits today.

Instead of looking at small-scale programs that were implemented long ago and never replicated, advocates of federal funding to expand early childhood education programs should examine the performance of Head Start, the federal government’s flagship early childhood education program.

Created as part of the War on Poverty in 1965, Head Start is a preschool grant program funded by the federal government. Head Start

KEY POINTS

- In an attempt to improve on the ineffective Head Start program, the Head Start CARES demonstration project tested three “evidence-based” interventions intended to help disadvantaged children to develop appropriate social-emotional behaviors.
- Experimental evaluations funded by the federal government found that Head Start CARES demonstration programs had little to no effect, compared with regular Head Start services, on the social-emotional and academic skills of participating children.
- Policymakers should take note that while the Head Start CARES demonstration was labeled as being “evidence-based,” the demonstration clearly revealed that such labeling does not mean that the program will be effective.
- The federal government has great difficulty demonstrating that it can replicate the results of successful small-scale social programs. Yet advocates of increased federal spending on early childhood education programs ignore the poor track record.

This paper, in its entirety, can be found at <http://report.heritage.org/bg3040>

The Heritage Foundation
214 Massachusetts Avenue, NE
Washington, DC 20002
(202) 546-4400 | heritage.org

Nothing written here is to be construed as necessarily reflecting the views of The Heritage Foundation or as an attempt to aid or hinder the passage of any bill before Congress.

is intended to provide a boost to disadvantaged children before they enter elementary school. Despite the program's long life, Head Start never underwent a scientifically rigorous evaluation of its effectiveness until Congress mandated a national-scale randomized impact evaluation in 1998. The Head Start Impact Study began in 2002.

The results—immediate-term, short-term, and long-term released in 2005, 2010, and 2012, respectively—are disappointing.³ Almost all of the benefits of participating in Head Start disappeared by kindergarten. Specifically, the evaluation found that the program largely failed to improve the cognitive, socio-emotional, health, and parenting outcomes of participating children in kindergarten, first grade, and third grade compared with the outcomes of similar children who did not participate.⁴

In search of evidence that Head Start can be an effective program, the Office of Planning, Research and Evaluation in the Administration for Children and Families in the U.S. Department of Health and Human Services (HHS) initiated the Head Start CARES (Classroom-based Approaches and Resources for Emotion and Social skill promotion) demonstration project in 2007. The demonstration tested

the effectiveness of three “enhancements” to regular Head Start services. Specifically, the demonstration assessed different methods of improving children's social-emotional development within the regular Head Start program, under the notion that preschool children from low-income families are more likely than their more well-off counterparts to be deficient in social, emotional, and behavioral development. Children who have greater difficulty regulating their emotions and behaviors may be less likely to receive appropriate instruction, to engage in positive learning behaviors, and to benefit from opportunities to learn from their peers.⁵

The Head Start CARES demonstration project tested three social-emotional interventions that were labeled as “evidence-based” to determine whether these interventions help disadvantaged children to develop appropriate social-emotional behaviors.⁶ The answer was “No.” Experimental evaluations released in 2014 and 2015 by the U.S. Department of Health and Human Services found that enhanced Head Start CARES demonstration programs had little to no effect, compared with regular Head Start services, on the social-emotional and academic skills of participating children.⁷ However,

1. Lawrence J. Schweinhart, Helen V. Barnes, and David P. Wiekart, *Significant Benefits: The High/Scope Perry Preschool Study Through Age 27* (Ypsilanti, MI: The High/Scope Press, 1993), and Frances A. Campbell and Craig T. Ramey, “Effects of Early Intervention on Intellectual and Academic Achievement: A Follow-Up Study of Children from Low-Income Families,” *Child Development*, Vol. 65, No. 2 (April 1994), pp. 684–698.
2. David B. Muhlhausen, “Do Federal Social Programs for Children Work?” testimony before the Committee on the Budget, U.S. Senate, June 26, 2013, <http://www.heritage.org/research/testimony/2013/06/do-federal-social-programs-for-children-work>.
3. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation, *Head Start Impact Study: First Year Findings*, June 2005, <http://www.acf.hhs.gov/programs/opre/resource/head-start-impact-study-first-year-findings> (accessed July 15, 2015); U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation, *Head Start Impact Study: Final Report*, January 2010, <http://www.acf.hhs.gov/programs/opre/resource/head-start-impact-study-final-report> (accessed July 15, 2015); and U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation, *Third Grade Follow-Up to the Head Start Impact Study: Final Report*, October 2012, <http://www.acf.hhs.gov/programs/opre/resource/third-grade-follow-up-to-the-head-start-impact-study-final-report> (accessed June 2, 2015).
4. For a review of the Head Start Impact Study, see David B. Muhlhausen, *Do Federal Social Programs Work?* (Santa Barbara, CA: Praeger, 2013); Lindsey M. Burke and David B. Muhlhausen, “Head Start Impact Evaluation Report Finally Released,” Heritage Foundation *Issue Brief* No. 3823, January 10, 2013, <http://www.heritage.org/research/reports/2013/01/head-start-impact-evaluation-report-finally-released>; and David B. Muhlhausen and Dan Lips, “Head Start Earns an F: No Lasting Impact for Children by First Grade,” Heritage Foundation *Background* No. 2363, January 21, 2010, <http://www.heritage.org/research/reports/2010/01/head-start-earns-an-f-no-lasting-impact-for-children-by-first-grade>.
5. Pamela Morris et al., *Impact Findings from the Head Start CARES Demonstration: National Evaluation of Three Approaches to Improving Preschoolers' Social and Emotional Competence*, U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation *OPRE Report* No. 2014-44, August 2014, pp. 1 and 3, <http://www.acf.hhs.gov/programs/opre/resource/impact-findings-from-the-head-start-cares-demonstration-national-evaluation-of-three-approaches-to-improving-preschoolers-social> (accessed June 2, 2015).
6. *Ibid.*, p. 3.
7. *Ibid.*, and JoAnn Hsueh et al., *Impacts of Social-Emotional Curricula on Three-Year-Olds: Exploratory Findings from the Head Start CARES Demonstration*, U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation, *OPRE Report* 2014-78, December 2014, <http://www.acf.hhs.gov/programs/opre/resource/exploratory-impacts-of-three-social-emotional-curricula-on-three-year-olds-in-the-head-start-cares-demonstration> (accessed July 15, 2015).

the demonstration programs were associated with beneficial outcomes on a minority of teachers' practices and classroom climate outcomes.⁸

While the demonstration focused on the effect of the enhancements on four-year-old children, in a separate report the evaluators assessed the effect of the enhancements on three-year-old children.⁹ The results were slightly different, but similarly poor. The authors caution that the findings for the three-year-old group should be considered exploratory because Head Start CARES was not explicitly designed for this age group.¹⁰

Misguided Belief in Early Childhood Education Programs

The Head Start CARES Demonstration was an attempt to improve the performance of the ineffective Head Start. By offering an intervention thought to help children to develop appropriate social-emotional behaviors, the program was expected to better equip participating children to learn than children who are deficient in this area.

While based on "evidence-based" interventions, the multisite random assignment evaluation found that, compared with regular Head Start services, enhanced interventions had little to no effect. Overall, "[n]one of the three enhancements had statistically significant impacts on measures of children's academic skills in kindergarten."¹¹ However, the demonstration programs were associated with beneficial outcomes on a minority of teachers' practices and classroom climate outcomes. For all three enhancements, the few small beneficial effects on child-level outcomes found in preschool disappeared during kindergarten. The pattern of initial effects quickly disappearing has been found to occur with regular Head Start and Early Head Start.¹²

Illusions and Good Intentions

Policymakers should take note that the Head Start CARES demonstration was labeled as being "evidence-based." Yet the demonstration clearly revealed that such labeling does not mean that the program will be effective.

The federal government has great difficulty demonstrating that it can successfully replicate the results of small-scale social programs originally thought to be successful. Yet advocates of increased federal spending on early childhood education programs ignore the federal government's poor track record in replicating small-scale programs.

Given the scientific uncertainty, advocates cannot answer the following question: Will increased federal spending on early childhood education programs improve children's futures? The evidence says probably not.

Advocates of expansion of such programs count on a positive response to a different question: Will proposing increased federal spending on early childhood programs make me feel better about myself and my good intentions toward children?

By creating the illusion that we are helping children in need, programs like Head Start and Early Head Start do a tremendous disservice by wasting both the resources and the political will for effective action. There may, in fact, be ways we can help children in need, but we will not find them if we believe, despite the evidence, that the right programs are already in place.

—**David B. Muhlhausen, PhD**, is a Research Fellow for Empirical Policy Analysis in the Center for Data Analysis, of the Institute for Economic Freedom and Opportunity, at The Heritage Foundation.

8. Morris et al., *Impact Findings from the Head Start CARES Demonstration*, and Hsueh et al., *Impacts of Social-Emotional Curricula on Three-Year-Olds*.

9. Hsueh et al., *Impacts of Social-Emotional Curricula on Three-Year-Olds*.

10. Ibid., p. 13.

11. Morris et al., *Impact Findings from the Head Start CARES Demonstration*, p. 102.

12. For a review of the multisite random assignment evaluations of Head Start and Early Head Start, see Muhlhausen, *Do Federal Social Programs Work?* and Muhlhausen, "Do Federal Social Programs for Children Work?"

Annex: Study Design and Detailed Results

The Head Start CARES demonstration project picked three social-emotional interventions to replicate because the HHS considered the interventions to be “evidence-based.”¹³ These interventions are:

1. The Incredible Years Teacher Training Program,
2. Preschool PATHS (Promoting Alternative Thinking Strategies), and
3. Tools of the Mind.¹⁴

The Incredible Years Teacher Training Program attempts to foster children’s ability to control their behavior by aiding teachers in maintaining an organized classroom. Preschool PATHS applies structured lessons to help children learn about emotions and acquire social problem-solving skills. Through structured make-believe play, Tools of the Mind attempts to promote children’s self-regulatory skills.¹⁵

“Evidence-based” programs are interventions that presumably have already been found to be effective based on randomized experiments in previous settings. However, merely replicating a program labeled as “evidence-based” does not necessarily mean that the same results will be produced.¹⁶ The most rigorous definitions of what qualifies as evidence-based require that a particular social program be found effective in more than one setting based on randomized experiments. However, HHS used a looser definition. Part of the criteria used by HHS for selection was the presence of “empirical evidence of the enhancement’s positive effect on

social-emotional outcomes, as reflected in at least one randomized controlled trial conducted on a sample of preschool, preferably low-income children.”¹⁷

The evaluations used to label the chosen interventions as evidence-based show varying degrees of success. For example, the two evaluations of The Incredible Years were noted for having statistically significant effect sizes ranging from 0.27 to 1.06, but did not have the same consistency in finding statistically meaningful results.¹⁸ While one of the evaluations found beneficial impacts that were statistically significant on four of six (66.7 percent) outcome measures,¹⁹ the other evaluation found that only six of the 35 (17.1 percent) outcome measures have statistically significant beneficial effects.²⁰

Head Start CARES Demonstration Methodology

The interventions provided by the three approaches are considered “enhancements” to regular Head Start services.²¹ With the control group receiving regular Head Start services, the impacts of the Head CARES demonstration should be interpreted as the effects of the enhancements over the existing Head Start services in the sites. In line with the traditional standards of social science, impacts (e.g., differences in outcomes between enhanced programs and regular Head Start) with *p*-values of 0.05 or less are considered to be statistically meaningful for this review.

According to the authors of the evaluation, “the comprehensive professional development supports helped ensure that each of the three enhancements

13. Morris et al., *Impact Findings from the Head Start CARES Demonstration*, p. 3.

14. *Ibid.*, p. 3.

15. *Ibid.*, p. iii.

16. Muhlhausen, *Do Federal Social Programs Work?* and Stuart M. Butler and David B. Muhlhausen, “Can Government Replicate Success?” *National Affairs*, No. 19 (Spring 2014), pp. 25–39, <http://www.nationalaffairs.com/publications/detail/can-government-replicate-success> (accessed April 13, 2015).

17. Morris et al., *Impact Findings from the Head Start CARES Demonstration*, p. 12.

18. *Ibid.*, p. 13.

19. C. Cybele Raver et al., “Targeting Children’s Behavior Problems in Preschool Classrooms: A Cluster-Randomized Controlled Trial,” *Journal of Consulting and Clinical Psychology*, Vol. 77, No. 2 (April 2007), p. 312, Table 3.

20. Pamela Morris et al., “Does a Preschool Social and Emotional Learning Intervention Pay Off for Classroom Instruction and Children’s Behavior and Academic Skills? Evidence from the Foundations of Learning Project,” *Early Education and Development*, Vol. 24, No. 7 (2013), p. 1032, Table 2, and p. 1034, Table 3.

21. Morris et al., *Impact Findings from the Head Start CARES Demonstration*, p. 3.

was delivered in Head Start classrooms with satisfactory fidelity.”²² In other words, the authors believe that each of the enhancements was implemented successfully.

For the demonstration, 17 Head Start sites were selected to be representative of the national population served by Head Start. The 17 sites were recruited in two groups.²³ For the four-year-old study, the first group consisted of four sites and participated in the demonstration for the 2009–2010 school year. The second group, comprised of 13 sites, participated during the 2010–2011 school year.

For the three-year-old study, the children attended mixed-age classrooms during the 2010–2011 school year. The mixed-age classrooms were composed of three-year-olds and four-year-olds, located in 56 Head Start centers within nine of the 17 grantees in the entire Head Start CARES sample.²⁴

Random Assignment. A drawback to the scientific rigor of the Head Start CARES demonstration is that the main unit of analysis—three-year-old and four-year-old children—were not randomly assigned to intervention and control groups. Instead, groups of four or eight similar Head Start centers under one grantee were randomly assigned to one of the three enhancement interventions or to a control group that conducted “business as usual.”²⁵ Therefore, the demonstration does not provide results that are as definitive as an evaluation that randomly assigned children to intervention and control groups. For the four-year-old cohort, 2,114 children participated in the demonstration, and 933 children were in the three-year-old cohort.²⁶ According to the evaluators, the demographics of the teachers and children assigned to the three intervention groups and control group were similar.²⁷

Statistical Significance. A “statistically significant” finding indicates that the effect of a particular intervention is statistically different from no effect. For example, if analysis finds that a social program has had a statistically significant effect on a particular outcome, then social scientists can conclude

with a high degree of confidence that the result was caused by the program, not by chance.

A “statistically insignificant” finding indicates that the effect of a particular intervention is no different from zero for statistical purposes. For example, if a social program is found to have a statistically insignificant effect on a particular outcome, the probability that chance caused the effect is too great for social scientists to conclude with confidence that the program produced the effect. In other words, the program had no statistically measurable effect on the particular outcome.

The common standard among social scientists for declaring a finding statistically significant is the 5 percent significance level ($p \leq 0.05$). This means that there is at least a 95 percent statistical probability that the program caused the effect and at most a 5 percent probability that the program had no measurable effect. Most social scientists use this rigorous standard of statistical significance because they want a high degree of confidence in their findings. Policymakers who make decisions based on social science research should also want a high degree of confidence. The 1 percent significance level ($p \leq 0.01$) is an even more rigorous standard, meaning that there is only a 1 percent probability that results were the product of chance.

Sometimes, social scientists will use the less rigorous standard of 10 percent ($p \leq 0.10$). Under this looser standard, social scientists are willing to risk a 10 percent chance of mistakenly concluding that the program had an effect, when it really had no effect at all. The 10 percent significance standard can be justified when social scientists are analyzing small samples, such as 100 cases. Studies using small sample sizes are less likely to be sensitive enough to find statistically significant findings at the 5 percent significance level than studies using much larger sample sizes.²⁸ Thus, social scientists sometimes use the less rigorous 10 percent significance level for small sample sizes. In contrast, the larger the sample size used in a study, the more sensitive the study will be

22. Ibid., p. 15.

23. Morris et al., *Impact Findings from the Head Start CARES Demonstration*, pp. 19 and 21.

24. Hsueh et al., *Impacts of Social-Emotional Curricula on Three-Year-Olds*, p. 15.

25. Morris et al., *Impact Findings from the Head Start CARES Demonstration*, p. ES-6.

26. Ibid., p. 27, Table 2.3, and Hsueh et al., *Impacts of Social-Emotional Curricula on Three-Year-Olds*, p. ES-9, Table ES-1.

27. Morris et al., *Impact Findings from the Head Start CARES Demonstration*, p. 28.

28. Mark W. Lipsey, *Design Sensitivity: Statistical Power for Experimental Research* (Newbury Park, CA: SAGE Publications, 1990).

in finding statistically significant effects. For this reason, most social scientists use the 5 percent confidence level when working with large sample sizes.

The Head Start CARES Demonstration provides results for classroom-level and child-level results for both cohorts. For this summary of the demonstration results, the findings for the classroom-level and child-level outcomes that are statistically significant at the 10 percent significance level are deemed “marginally” statistically significant to reflect the lesser degree of confidence in the findings. However, due to the relatively smaller sample size for the classroom-level results, impacts that are statistically significant at the 10 percent level should be considered more statistically meaningful than the child-level results that are statistically significant at the 10 percent level.

Effect Sizes. The Head Start CARES Demonstration reports present the findings using standardized effect sizes based on differences between the mean scores of the intervention and control groups. Standardized effect sizes allow for the comparison of multiple outcomes that have varying scales of measurement. After adjusting for covariates, the mean outcome for the control group was subtracted from the mean outcome for the intervention group. The difference was divided by the standard deviation for the control group. The calculation used for this effect size (*ES*) is called Glass’s Δ .

For example, an effect size of 0.50 using Glass’s Δ would signify that the mean score for the intervention group on a particular outcome is half a standard deviation above the mean score for the control group. Alternatively, a Glass’s Δ of –1.0 would indicate that the mean score for the intervention group is one standard deviation below the mean for the control group.

Social scientists have debated the merits of rules of thumb for interpreting effect sizes. The first rule of thumb for classifying the magnitude of effect sizes for behavioral science was proposed by psychologist

Jacob Cohen in 1977.²⁹ Based on Cohen’s review of behavioral science research, he proposed the following benchmarks for interpreting the magnitude of standardized mean differences effect sizes:

Small	$ES \leq 0.20$
Medium	$ES = 0.50$
Large	$ES \geq 0.80$

For example, an effect size of 0.15 or –0.15 would be considered small, while an effect size of 1.0 or –1.0 would be considered large. Cohen’s benchmark has been criticized for not being based on a systematic analysis, but on his generalization of the research literature.³⁰ The authors of the Head Start CARES Demonstration report for the four-year-old cohort recommend a two-tier set of benchmarks for judging effect sizes. For the classroom-level outcomes, they consider moderate and large effects to be around effect sizes of 0.50 and 0.80, respectively. Presumably, small effects are judged to be around 0.20. However, for the child-level outcomes, they recommend³¹ the following benchmarks:

Small	$ES \leq 0.20$
Medium	$0.20 \leq ES \leq 0.40$
Large	$ES > 0.40$

This recommendation for the child-level outcomes drastically lowers the bar for judging which effect sizes fall within the medium and large classifications. The authors offered the following justification for lowering the standard: “Given that effects on children must occur as a result of changes in teachers’ practices, effects were expected to be smaller on child outcomes than on teachers’ practices.”³²

Table 1 presents benchmarks for judging the magnitude of effect sizes suggested by Mark W. Lipsey of Vanderbilt University.³³ These benchmarks are based on an analysis of selected mean effect sizes from 186 meta-analyses of psychological, educational, and behavioral treatment programs.³⁴

29. Jacob Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Hillsdale, NJ: Lawrence Erlbaum Associates, 1998), pp. 25–26.

30. Howard S. Bloom et al., “Performance Trajectories and Performance Gaps as Achievement Effect-Size Benchmarks for Educational Interventions,” MDRC *Working Papers on Research Methodology*, October 2008, <http://www.mdrc.org/publication/performance-trajectories-and-performance-gaps-achievement-effect-size-benchmarks> (accessed July 15, 2015).

31. Morris et al., *Impact Findings from the Head Start CARES Demonstration*, p. 51, Box 3.6.

32. Ibid.

33. Lipsey, *Design Sensitivity*, p. 56, Table 3.5.


34. Ibid., p. 50.

TABLE 1

Small, Medium, and Large Benchmarks for Effect Sizes

Range	Values of Effect Size	Midpoint
Small (lower 33%)	0.00 to 0.32	0.15
Medium (middle 34%)	0.33 to 0.55	0.45
Large (upper 33%)	0.56 to 1.20	0.90

Source: Mark W. Lipsey, *Design Sensitivity: Statistical Power for Experimental Research* (Newbury Park, CA: SAGE Publications, 1990), p. 56, Table 3.5.

BG 3040  heritage.org

After taking the absolute values of the effect sizes, Lipsey’s benchmarks are based on dividing a range of effect sizes into thirds. The first-third constitutes effect sizes that are considered “small” that range from 0.00 to 0.32 with a midpoint of 0.15. The middle third constitutes effect sizes that are considered “medium” that range from 0.33 to 0.55 with a midpoint of 0.45. The last third constitutes effect sizes that are considered “large” that range from 0.56 to 1.20 with a midpoint of 0.90. Lipsey’s and Cohen’s benchmarks are surprisingly similar.

For the purpose of this review, the benchmarks in Table 1 will be used to summarize the results of the Head Start CARES Demonstration. The decision to use these benchmarks is based on the need to assess the effectiveness of the Head Start Cares Demonstration relative to the effectiveness of other psychological, educational, and behavioral interventions. The use of the benchmarks in Table 1 avoids the problem of scaling down what is considered effective because the program being assessed has relatively smaller effects compared with other psychological, educational, and behavioral treatments.

The Incredible Years: Impact Summary of Four-Year-Old Cohort Results

Table 2 summarizes the findings for four-year-olds participating in The Incredible Years. Overall, The Incredible Years failed to have statistically significant effects on the majority of outcomes. Only seven of 31 classroom-level and nine of 33 child-level

outcomes were statistically significant. The statistically significant effect sizes for the classroom-level outcomes were moderate, while the effect sizes of the statistically significant child-level outcomes were small.

Classroom-Level Impacts. Overall, The Incredible Years failed to produce statistically meaningful impacts on the majority of classroom-level impacts.

Teachers’ Practices. The Incredible Years classrooms had statistically significant beneficial impacts on five of 17 (29.4 percent) teachers’ practices outcomes that assessed classroom management, social-emotional instruction, and scaffolding.³⁵ Each of the 17 teachers’ practices outcomes is based on a five-point scale with a one designated as “low” and a five designated as “high.” For classroom management, intervention group teachers were observed managing their classrooms better than their counterparts on consistency/routine ($ES = 0.44, p = 0.05$), positive behavior management ($ES = 0.55, p = 0.01$), negative behavior management ($ES = -0.32, p = 0.05$), and attention/engagement ($ES = 0.53, p = 0.01$). The intervention had no effect on consistency/routine, preparedness, and classroom awareness.

For social-emotional instruction, intervention group teachers displayed a greater level of social problem solving ($ES = 0.40, p = 0.05$). Otherwise, they did not appear to display statistically significant differences at traditional levels. The impacts for overall social-emotional instruction ($ES = 0.30, p = 0.10$), emotion modeling ($ES = 0.38, p = 0.10$), and social awareness ($ES = 0.40, p = 0.10$) were marginally statistically significant. For emotion expression, emotion regulation, and the provision of interpersonal support, differences between teacher practices in the intervention and controls were statistically indistinguishable from zero. The implementation of The Incredible Years failed to affect all three outcomes assessing the teachers’ observed performance engaged in scaffolding—the act of aiding a child to achieve a challenging task or obtain a skill that is just beyond the child’s current capability.

Classroom Climate. The classroom-level impacts on classroom climate at the time of the preschool follow-up for the four-year-old cohort generally found the intervention to be ineffective. Each of the 14 classroom climate outcomes is based on a seven-point scale with a one designated as “low” and a

35. Morris et al., *Impact Findings from the Head Start CARES Demonstration*, p. 57, Table 4.1.

TABLE 2

The Incredible Years: Effect Size Findings for 4-Year-Old Cohort

KEY

- — No effect
- — Small effect (0.00 to 0.32)
- — Medium effect (0.33 to 0.55)
- — Large effect (0.56 to 1.20)

		CLASSROOM-LEVEL IMPACTS		CHILD-LEVEL IMPACTS					
		Teachers' Practices	Classroom Climate	PRESCHOOL FOLLOW-UP			KINDERGARTEN FOLLOW-UP		
				Execution Function, Behavior Regulation, and Learning Behaviors	Socio-Emotional Skills and Social Behaviors	Early Verbal, Literacy, and Match Skills	Behavior and Social Skills	Academic Skills	Grade Retention and Special Education Services
BENEFICIAL	Highly Significant (P≤0.01)	●●			○				○
	Significant (P≤0.05)	●●●○	●●	○	○○○	○○○			
	Marginally Significant (P≤0.10)	●●●○			○		○		
NO EFFECT	(P>0.10)	●●●●●●●●	●●●●●●●●●●	●●●●●●●●	●	●●●●	●●●●●●●●	●●●●	●●
HARMFUL	Marginally Significant (P≤0.10)								
	Significant (P≤0.05)								
	Highly Significant (P≤0.01)								

Note: Effect sizes are presented for impacts that are marginally statistically significant ($p \leq 0.10$) and statistically significant ($p \leq 0.05$ and $p \leq 0.01$).
Source: Pamela Morris et al., *Impact Findings from the Head Start CARES Demonstration: National Evaluation of Three Approaches to Improving Preschoolers' Social and Emotional Competence*, U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation, August 2014, p. 57, Table 4.1, p. 59, Table 4.2, p. 61, Table 4.3, p. 63, Table 4.4, p. 93, Table 7.1, pp. 100–101, Table 8.1, p. 103, Table 8.2, and p. 105, Table 8.3, <http://www.mdrc.org/publication/impact-findings-head-start-cares-demonstration> (accessed June 17, 2015).

BG 3040  heritage.org

seven designated as “high.” For classroom climate, only two of 14 (14.3 percent) outcomes yielded statistically significant results.³⁶ The Incredible Years classrooms were perceived to have lower negative climates ($ES = -0.26, p = 0.05$) and better displays of behavioral management ($ES = 0.39, p = 0.05$), compared with traditional Head Start classrooms. Otherwise, The Incredible Years classrooms did not

display any statistically meaningful differences on 12 other measures of emotional support, classroom organization, instructional support, and literacy focus.

Child-Level Impacts for Preschool Follow-Up. On the whole, The Incredible Years failed to have statistically significant impacts on the majority of child-level preschool outcomes.

36. Ibid., p. 59, Table 4.2.

Executive Function, Behavior Regulation, and Learning Behaviors. The Incredible Years has a small statistically significant effect in only one of seven (14.3 percent) child-level measures of executive function, behavior regulation, and learning behaviors at the time of the preschool follow-up for the four-year-old cohort.³⁷ While The Incredible Years had no effect on two measures of executive function and four measures of teacher-reported behavior problem assessments, the intervention did yield small beneficial results for teacher-reported work-related skills ($ES = 0.17, p = 0.05$).

Socio-emotional Skills and Social Behaviors. While small in magnitude, The Incredible Years did produce statistically significant effects on child-level social-emotional skills and social behaviors at the time of the preschool follow-up for the four-year-old cohort for four of six (66.7 percent) outcomes.³⁸ Participation in The Incredible Years had statistically significant beneficial impacts on facial emotions identification ($ES = 0.13, p = 0.05$), Challenging Situations competent response ($ES = 0.14, p = 0.05$), Challenging Situations aggressive response ($ES = -0.14, p = 0.05$), and teacher-reported Social Skills Rating Scale ($ES = 0.28, p = 0.01$). The effect for emotions situations identification was marginally statistically significant ($ES = 0.10, p = 0.10$), while the effect of the program on teacher-reported interpersonal skills was statistically indistinguishable from zero.

Early Verbal, Literacy, and Math Skills. While the three Head Start CARES enhancements focused on children's social-emotional skills and behaviors, the demonstration also assessed the programs' impact on pre-academic skills assessed during the Head Start program year. The authors of the four-year-old cohort study propose that "it is also possible that supporting children's social-emotional skills and behaviors, with related benefits to children's learning behaviors, could translate into improved pre-academic outcomes in school." However, the authors conclude: "These exploratory analyses yield no

consistent evidence that any of the enhancements led to improved pre-academic skills in the Head Start year."³⁹

For three of the six (50 percent) early verbal, literacy, and math skills assessments, The Incredible Years was associated with beneficial impacts for the participating children.⁴⁰ For three standardized measures of pre-academic skills (Woodcock-Johnson Letter-Word Identification, Woodcock-Johnson Applied Problems, and Expressive One-Word Picture Vocabulary Test), children in The Incredible Years intervention group failed to display statistically meaningful differences when compared with children in regular Head Start. However, based on teacher reports, children in the intervention group had higher scores on pre-academic skills, including general knowledge ($ES = 0.29, p = 0.05$), language and literacy ($ES = 0.27, p = 0.05$), and mathematical thinking ($ES = 0.32, p = 0.05$) that were statistically significant. Accordingly, the authors conclude that "these findings should be interpreted cautiously, given the lack of convergence in findings between the standardized assessments and teachers' reports."⁴¹

In sum, The Incredible Years failed to produce statistically meaningful results for pre-academic skills based on standardized tests for the four-year-old cohort during the preschool follow-up.

Child-Level Impacts for Kindergarten Follow-Up. The kindergarten follow-up attempts to ascertain whether the three preschool enhancements have impacts one year later. The outcomes assessed are limited to teacher-reported assessments, which are not corroborated with standardized tests. Overall, "[t]here was little evidence that any of the three enhancements had sustained impacts into kindergarten, based on the limited information collected."⁴²

Behavior and Social Skills. All seven assessments of behavior and social skills in kindergarten failed to yield statistically significant differences between children in The Incredible Years and regular Head Start classrooms.⁴³ However, for the single measure

37. Ibid., p. 61, Table 4.3.

38. Ibid., p. 63, Table 4.4.

39. Ibid., p. 91.

40. Ibid., p. 93, Table 7.1.

41. Ibid., p. 92.

42. Ibid., p. 97.

43. Ibid., pp. 100-101, Table 8.1.

of the behavior problem of externalizing, there was a small, but marginally statistically significant beneficial impact ($ES = -0.13$, $p = 0.10$). Despite this uncertain finding, this enhancement had no effect on reducing behavior problems, improving learning behaviors as reported by teachers, or improving social behaviors as reported by teachers and parents. Overall, The Incredible Years failed to affect the behavior and social skills of children in kindergarten on all eight measures.

Academic Skills. Teachers assessed their students on academic skills relating to general knowledge, language and literacy, and mathematical thinking. On all three of these kindergarten teacher-reported measures, The Incredible Years had no statistically measurable effects.⁴⁴

Grade Retention and Special Education Services. As reported by teachers, The Incredible Years had no discernable effect on two of three (66.7 percent) outcome assessments of expected grade retention and the receipt of special education services of children in kindergarten.⁴⁵ Participation in The Incredible Years has no effect on the expectation of a child being retained or held back in kindergarten for an additional year or the receipt of special education services as reported by teachers. However, parents reported a small and statistically significant effect of the program on their children being more likely to receive special education services ($ES = 0.19$, $p = 0.05$). According to the authors,

This impact on special education is plausible, since The Incredible Years may have made teachers more likely to identify serious behavior problems and therefore refer children to special services when they entered kindergarten. While evaluations in elementary school often examine impacts on the use of special education because of its cost implications for the school system, increases in the use of these services in kindergarten might bode well if it meant that children's problems were being identified early.⁴⁶

Thus, the authors interpret as a beneficial outcome the increased likelihood of children being

placed in special education for behavior problems displayed in kindergarten after participating in The Incredible Years preschool program. While this interpretation may have credence, an alternative explanation may be just as likely. Because The Incredible Years children were placed in special education during kindergarten and after graduating from the preschool intervention, the finding could be plausibly interpreted as a harmful impact. In other words, participation in The Incredible Years may have resulted in children displaying behavior problems that would lead them to being placed in special education in kindergarten.

Executive Function, Behavior Regulation, and Learning Behaviors. For the preschool follow-up that assessed child-level impacts, The Incredible Years had statistically significant beneficial impacts on one of seven (14.3 percent) executive function, behavior regulation, and learning behaviors outcomes, four of six (66.7 percent) socio-emotional skills and social behaviors outcomes, and three of six (50 percent) early verbal, literacy, and math skills outcomes. Afterward, the benefits of The Incredible Years diminished. For the kindergarten follow-up, The Incredible Years had statistically significant beneficial impacts on zero of eight (0 percent) behavior and social skills outcomes, zero of three (0 percent) academic skills outcomes, and one of three (33.3 percent) on grade retention and special education services assessments.

PATHS: Impact Summary of Four-Year-Old Cohort Results

The findings for four-year-olds participating in PATHS are summarized in Table 3. Overall, 11 of 31 (35.5 percent) classroom-level and six of 33 (18.2 percent) child-level outcomes were statistically significant. The statistically significant effect sizes for the classroom-level outcomes ranged from small to large impacts, while all the effect sizes for the child-level outcomes were small.

Classroom-Level Impacts. Similar to The Incredible Years, the effects of PATHS diminished over time. For the kindergarten follow-up that assessed child-level impacts, PATHS had

44. Ibid., p. 103, Table 8.2.

45. Ibid., p. 105, Table 8.3.

46. Ibid., p. 104.

TABLE 3

PATHS: Effect Size Findings for 4-Year-Old Cohort

KEY

- —No effect
- —Small effect (0.00 to 0.32)
- ◐ —Medium effect (0.33 to 0.55)
- ◑ —Large effect (0.56 to 1.20)

		CLASSROOM-LEVEL IMPACTS		CHILD-LEVEL IMPACTS					
		Teachers' Practices	Classroom Climate	PRESCHOOL FOLLOW-UP			KINDERGARTEN FOLLOW-UP		
				Execution Function, Behavior Regulation, and Learning Behaviors	Socio-Emotional Skills and Social Behaviors	Early Verbal, Literacy, and Match Skills	Behavior and Social Skills	Academic Skills	Grade Retention and Special Education Services
BENEFICIAL	Highly Significant (P≤0.01)	<div><div></div><div></div><div></div><div></div><div></div></div>			<div><div></div><div></div></div>				<div><div></div></div>
	Significant (P≤0.05)	<div><div></div><div></div></div>	<div><div></div><div></div><div></div></div>	<div><div></div></div>	<div><div></div><div></div></div>				
	Marginally Significant (P≤0.10)								
NO EFFECT	(P>0.10)	<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>	<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>	<div><div></div><div></div><div></div><div></div><div></div><div></div></div>	<div><div></div><div></div></div>	<div><div></div><div></div><div></div><div></div><div></div><div></div></div>	<div><div></div><div></div><div></div><div></div><div></div><div></div></div>	<div><div></div><div></div><div></div></div>	<div><div></div><div></div></div>
HARMFUL	Marginally Significant (P≤0.10)								
	Significant (P≤0.05)								
	Highly Significant (P≤0.01)								

Note: Effect sizes are presented for impacts that are marginally statistically significant ($p \leq 0.10$) and statistically significant ($p \leq 0.05$ and $p \leq 0.01$).
Source: Pamela Morris et al., *Impact Findings from the Head Start CARES Demonstration: National Evaluation of Three Approaches to Improving Preschoolers' Social and Emotional Competence*, U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation, August 2014, p. 71, Table 5.1, p. 72, Table 5.2, p. 75, Table 5.3, p. 77, Table 5.4, p. 94, Table 7.2, pp. 100–101, Table 8.1, p. 103, Table 8.2, and p. 105, Table 8.3, <http://www.mdrc.org/publication/impact-findings-head-start-cares-demonstration> (accessed June 17, 2015).

BG 3040  heritage.org

statistically significant beneficial impacts on zero of eight (0 percent) behavior and social skills outcomes, zero of three (0 percent) academic skills outcomes, and one of three (33.3 percent) on grade retention and special education services assessments.

Teachers' Practices. PATHS had moderate to large effects that were statistically significant on eight of

the 17 (47.1 percent) teachers' practices outcomes.⁴⁷ For classroom management, teachers trained in PATHS displayed a higher degree of positive behavior management than their counterparts providing traditional Head Start services ($ES = 0.33$, $p = 0.05$). However, for assessments of overall classroom management, consistency/routine, preparedness,

47. Ibid., p. 71, Table 5.1.

classroom awareness, negative behavior management, and attention/engagement, differences between PATHS and traditional Head Start teachers were statistically indistinguishable from zero.

In contrast to classroom management, PATHS teachers were consistently assessed to provide higher levels of social-emotional instruction, compared with regular Head Start teachers.⁴⁸ Teachers trained in PATHS displayed higher degrees of overall social-emotional instruction ($ES = 0.92, p = 0.01$), emotion modelling ($ES = 1.36, p = 0.01$), emotion expression ($ES = 0.82, p = 0.01$), emotion regulation ($ES = 0.58, p = 0.01$), social awareness ($ES = 0.92, p = 0.01$), social problem solving ($ES = 0.82, p = 0.01$), and provision of interpersonal support ($ES = 0.34, p = 0.05$).

The implementation of PATHS failed to affect the three outcomes assessing the teachers' observed performance engaged in scaffolding—the act of aiding a child to achieve a challenging task or obtain a skill that is just beyond the child's current capability.⁴⁹

Classroom Climate. PATHS had small to moderate effects that were statistically significant on three of the 14 (21.4 percent) measures of classroom climate.⁵⁰ Compared with regular Head Start classrooms, implementation of PATHS was found to have no effect on overall emotional support, positive climate, negative climate, teacher sensitivity, and regard for student perspectives. This trend also prevailed for the four classroom organization outcomes measures: overall classroom organization, behavior management, productivity, and instructional learning formats.

In contrast, the implementation of PATHS was associated with improved instructional support ($ES = 0.27, p = 0.05$), concept development ($ES = 0.33, p = 0.05$), and quality of feedback ($ES = 0.29, p = 0.05$) in the classrooms.⁵¹ The intervention had no effect on language modelling and literacy focus in the classrooms.

Child-Level Impacts for Preschool Follow-Up. PATHS largely failed to produce statistically

meaningful child-level impacts at the preschool follow-up.

Executive Function, Behavior Regulation, and Learning Behaviors. PATHS had a small effect on one of seven (14.3 percent) measures of executive function, behavior regulation, and learning behaviors of children during the preschool follow-up.⁵² While the intervention failed to produce statistically significant effects on all measures of executive function (Head-to-Toes and Pencil Tap) and teacher-reported behavior problems (overall behavior problems, externalizing, hyperactivity, and internalizing), participation in the programs did produce a small beneficial effect on teacher-reported work-related skills of children ($ES = 0.20, p = 0.05$).

Socio-emotional Skills and Social Behaviors. Participation in PATHS produced small statistically meaningful impacts on four of six (66.7 percent) measures of emotional knowledge, social problem solving, and social behaviors.⁵³ For emotional knowledge, children in the intervention group displayed higher levels of facial emotions identification ($ES = 0.29, p = 0.01$) and emotions situations identification ($ES = 0.23, p = 0.01$), compared with the children in the control group.

For social problem solving, intervention group children were rated slightly higher on Challenging Situations competent response assessment ($ES = 0.17, p = 0.05$), while participation in the program failed to affect the Challenging Situations aggressive response assessment.⁵⁴ A similar pattern held for the social behaviors assessed by teacher reports. Children in the intervention group were reported by teachers to have higher scores on the Social Skills Rating Scale, compared with children in the control group ($ES = 0.19, p = 0.05$). However, participation in PATHS failed to produce a statistically meaningful impact on the interpersonal skills of the children as reported by teachers.

Early Verbal, Literacy, and Math Skills. For PATHS, the intervention failed to produce statistically

48. Ibid.

49. Ibid.

50. Ibid., p. 72, Table 5.2.

51. Ibid.

52. Ibid., p. 77, Table 5.4.

53. Ibid., p. 75, Table 5.3.

54. Ibid.

meaningful results for all six measures of pre-academic skills for the four-year-old cohort during the preschool follow-up.⁵⁵ For pre-academic skills, the children in the PATHS classrooms did not display any statistically meaningful differences on the Woodcock-Johnson Letter-Word Identification, Woodcock-Johnson Applied Problems, and Expressive One-Word Picture Vocabulary Test. The same held true for teacher-reported assessments of general knowledge, language and literacy, and mathematical thinking.

Child-Level Impacts for Kindergarten Follow-Up. On nearly all of the child-level outcomes for the kindergarten follow-up, PATHS failed to produce statistically meaningful impacts, compared with regular Head Start.

Behavior and Social Skills. On all eight outcome measures of behavior regulation and social behaviors reported by teachers and parents, children in the PATHS intervention group failed to yield statistically measurable impacts, compared with their counterparts in the regular Head Start classrooms.⁵⁶

Academic Skills. On all three kindergarten teacher-reported measures of academic skills (general knowledge, language and literacy, and mathematical thinking), PATHS failed to yield statistically measurable effects.⁵⁷

Grade Retention and Special Education Services. PATHS had a small statistically significant effect on only one of three (33.3 percent) measures of grade retention and receipt of special education services.⁵⁸ As reported by teachers, children in the PATHS intervention group were less likely to be expected to be retained in kindergarten for an additional year, compared with their counterparts in the Head Start control group ($ES = 0.24, p = 0.01$).⁵⁹ For the teacher-reported and parent-reported outcomes for participation in special education services, PATHS appears to have no statistically measurable effect.

Tools of the Mind: Impact Summary of Four-Year-Old Cohort Results

The findings for four-year-olds participating in Tools of the Mind are summarized in Table 4. Overall, four of 31 (12.9 percent) classroom-level and two of 33 (6.1 percent) child-level outcomes were statistically significant. The statistically significant effect sizes for the classroom-level outcomes range from small to medium impacts, while the statistically significant effect sizes for the child-level outcomes were small.

Classroom-Level Impacts. Overall, Tools of the Mind failed to produce statistically significant impacts on classroom outcomes.

Teachers' Practices. For three of 17 (17.6 percent) measures of teachers' practices, Tools of the Mind was associated with small, yet statistically significant effects.⁶⁰ On all 14 measures of classroom management and social-emotional instruction, the practices of teachers in the Tools of the Mind intervention and control groups did not have statistically meaningful differences.⁶¹ However, compared with teachers in the control group, teachers in the Tools of the Mind intervention group were reported to be more engaged in all three scaffolding measures: overall scaffolding ($ES = 0.68, p = 0.01$), dramatic play ($ES = 0.66, p = 0.01$), and peer interaction ($ES = 0.57, p = 0.01$).

Classroom Climate. On one of 14 (7.1 percent) measures of classroom climate, Tools of the Mind classrooms had a better rating than the regular Head Start as usual classrooms.⁶² Tools of the Mind did appear to have a moderate effect on improving the teachers' use of literacy strategies in the classrooms ($ES = 0.50, p = 0.01$). However, Tools of the Mind appears to have had no effect on 13 measures of emotional support, classroom organization, and instructional support.

Child-Level Impacts for Preschool Follow-Up. With a few exceptions, Tools of the Mind, compared with regular Head Start, failed to yield statistically

55. Ibid., p. 94, Table 7.2.

56. Ibid., pp. 100-101, Table 8.1.

57. Ibid., p. 103, Table 8.2.

58. Ibid., p. 105, Table 8.3.

59. Ibid.

60. Ibid., p. 83, Table 6.1.

61. Ibid.

62. Ibid., p. 85, Table 6.2.

TABLE 4

Tools of the Mind: Effect Size Findings for 4-Year-Old Cohort

KEY

- —No effect
- —Small effect (0.00 to 0.32)
- —Medium effect (0.33 to 0.55)
- —Large effect (0.56 to 1.20)

		CLASSROOM-LEVEL IMPACTS		CHILD-LEVEL IMPACTS					
		Teachers' Practices	Classroom Climate	PRESCHOOL FOLLOW-UP			KINDERGARTEN FOLLOW-UP		
				Execution Function, Behavior Regulation, and Learning Behaviors	Socio-Emotional Skills and Social Behaviors	Early Verbal, Literacy, and Match Skills	Behavior and Social Skills	Academic Skills	Grade Retention and Special Education Services
BENEFICIAL	Highly Significant (P≤0.01)	●●●●	●●						
	Significant (P≤0.05)				○○				
	Marginally Significant (P≤0.10)					○			
NO EFFECT	(P>0.10)	●●●●●●●●	●●●●●●●●	●●●●●●●●	●●●●●●	●●●●●●●●	●●●●●●●●	●●●●	●●●●
HARMFUL	Marginally Significant (P≤0.10)								
	Significant (P≤0.05)								
	Highly Significant (P≤0.01)								

Note: Effect sizes are presented for impacts that are marginally statistically significant (p≤0.10) and statistically significant (p≤0.05 and p≤0.01).
Source: Pamela Morris et al., *Impact Findings from the Head Start CARES Demonstration: National Evaluation of Three Approaches to Improving Preschoolers' Social and Emotional Competence*, U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation, p. 83, Table 6.1, p. 85, Table 6.2, p. 87, Table 6.3, p. 88, Table 6.4, p. 95, Table 7.3, pp. 100–101, Table 8.1, p. 103, Table 8.2, and p. 105, Table 8.3, <http://www.mdrc.org/publication/impact-findings-head-start-cares-demonstration> (accessed June 17, 2015).

BG 3040  heritage.org

significant results for child-level outcomes during the preschool follow-up.

Executive Function, Behavior Regulation, and Learning Behaviors. Tools of the Mind failed to yield statistically meaningful impacts on all seven outcome measures for executive function, behavior regulation, and learning behaviors of children.⁶³

Socio-emotional Skills and Social Behaviors. For two of six (33.3 percent) measures of socio-emotional skills and social behaviors during preschool, Tools of the Mind has small, but statistically significant impacts.⁶⁴ For both measures of emotional knowledge, children participating in Tools of the Mind classrooms displayed slightly higher scores

63. Ibid., p. 87, Table 6.3.

64. Ibid., p. 88, Table 6.4.

on facial emotions identification ($ES = 0.12, p = 0.05$) and emotions situations identification ($ES = 0.13, p = 0.05$), compared with their counterparts in the regular Head Start classrooms. Otherwise, children in Tools of the Mind classrooms failed to display differences compared with members of the control group on four measures of social problem solving and social behaviors. In sum, Tools of the Mind failed to produce statistically meaningful results for four of six (66.7 percent) child-level impacts on social-emotional skills and social behaviors.

Early Verbal, Literacy, and Math Skills. Similar to PATHS, the Tools of the Mind intervention failed to produce statistically meaningful results for all six measures of pre-academic skills for the four-year-old cohort during the preschool follow-up.⁶⁵ However, the extremely small beneficial finding for the Woodcock-Johnson Applied Problems assessment ($ES = 0.09, p = 0.10$) was marginally statistically significant.

Child-Level Impacts for Kindergarten Follow-Up. Without exception, Tools of the Mind failed to produce statistically meaningful impacts on the child-level outcomes during the kindergarten follow-up.

Behavior and Social Skills. Similarly to the findings for the other enhancements, Tools of the Mind failed to yield statistically measurable impacts on all eight outcome measures of behavior regulation and social behaviors as reported by teachers and parents.⁶⁶

Academic Skills. On all three of these kindergarten teacher-reported measures (general knowledge, language and literacy, and mathematical thinking), Tools of the Mind had no statistically measurable effects.⁶⁷

Grade Retention and Special Education Services. For all measures of expected grade retention and special education services participation, Tools of the Mind failed to yield statistically measurable impacts compared with regular Head Start services.⁶⁸

Three-Year-Old Impact Findings

The outcome measures to assess the effectiveness of the enhancements for the three-year-old group are not the same as the outcomes used for the four-year-old group. In particular, direct assessments through standardized testing instruments were not used, so teacher-reported assessments are the sole method used to assess social, emotional, and behavioral competencies.⁶⁹ Nevertheless, “exclusive reliance on teacher reports can be a limitation; teachers’ ratings may be influenced by their own perceptions, and teachers who were trained in the Head Start CARES enhancements might perceive children’s behavior differently from those who did not receive this training, regardless of whether children’s *actual* behaviors changed.”⁷⁰ Further, the authors caution against comparing the three-year-old results with the four-year-old results because the samples vary in enough ways that impede drawing conclusions about why the results differ.⁷¹ For instance,

the sample of mixed-age classrooms differed from the full sample of classrooms on a number of characteristics, including the classrooms’ locations around the country, the types of organizations in which the grantees were located, and baseline levels of classroom climate. Therefore, results for the full sample of classrooms should not be used to make head-to-head comparisons with impacts on outcomes in the sample of mixed-age classrooms.⁷²

Due to the small sample sizes of the mixed-age classrooms and three-year-old cohort, only the results of the three enhancements pooled together are presented in this summary report.⁷³ The pooled results test “whether any of the enhancements affects class-level and child outcomes and maximizes statistical power.”⁷⁴ The results are presented in Table 5.

65. Ibid., p. 95, Table 7.3.

66. Ibid., pp. 100–101, Table 8.1.

67. Ibid., p. 103, Table 8.2.

68. Ibid., p. 105, Table 8.3.

69. Hsueh et al., *Impacts of Social-Emotional Curricula on Three-Year-Olds*, p. 13.

70. Ibid., p. ES-6 (emphasis in the original).

71. Ibid., p. 27.

72. Ibid., p. 27.

73. For the individual findings for each of the enhancements, see *ibid.*

74. Ibid., p. 14.

TABLE 5

Pooled Enhanced: Effect Size Findings for 3-Year-Old Cohort

KEY

- —No effect
- —Small effect (0.00 to 0.32)
- ◐ —Medium effect (0.33 to 0.55)
- ◑ —Large effect (0.56 to 1.20)

Note: Effect sizes are presented for impacts that are marginally statistically significant ($p \leq 0.10$) and statistically significant ($p \leq 0.05$ and $p \leq 0.01$).

Source: JoAnn Hsueh et al., *Impacts of Social-Emotional Curricula on Three-Year-Olds: Exploratory Findings from the Head Start CARES Demonstration*, U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation, December 2014, p. 26, Table 3, p. 28, Table 4, p. 38, Table 7, and p. 45, Table 9, <http://www.acf.hhs.gov/programs/opre/resource/exploratory-impacts-of-three-social-emotional-curricula-on-three-year-olds-in-the-head-start-cares-demonstration> (accessed June 17, 2015).

		MIXED-AGE CLASSROOM-LEVEL IMPACTS		INDIVIDUAL-LEVEL PRESCHOOL FOLLOW-UP	
		Teachers' Practices	Classroom Climate	Social-Emotional Outcomes	Pre-Academic Skills
BENEFICIAL	Highly Significant ($P \leq 0.01$)			○	
	Significant ($P \leq 0.05$)	◐		○○○	
	Marginally Significant ($P \leq 0.10$)	◐◐		○	
NO EFFECT	($P > 0.10$)	●●●●●●●●●●●●●●●●	●●●●●●●●●●●●●●●●	●●●●●●●●●●●●●●●●	●●●●●●●●●●●●●●●●
HARMFUL	Marginally Significant ($P \leq 0.10$)				
	Significant ($P \leq 0.05$)				
	Highly Significant ($P \leq 0.01$)				

BG 3040  heritage.org

Mixed-Age Classroom-Level Impacts on Teacher Practices at Preschool Follow-Up. The effect of the three enhancements on teachers' practices during preschool was statistically significant on one of the 17 (5.9 percent) outcome measures.⁷⁵ Teachers in the enhanced classroom were observed to engage in moderately higher social problem solving ($ES = 0.40$, $p = 0.05$), compared with their counterparts in the regular Head Start classrooms. For two measures, teachers in enhanced classrooms had beneficial impacts that were marginally statistically significant on emotion modelling ($ES = 0.39$, $p = 0.10$) and social awareness ($ES = 0.40$, $p = 0.10$).

On all seven measures of teacher practices, the teachers in the pooled enhancement classrooms were not observed to manage their classrooms differently compared with their counterparts in the

regular Head Start classrooms.⁷⁶ For three scaffolding measures, the effects of pooled enhancements failed to produce statistically meaningful differences from the effect of Head Start as usual.

Classroom-Level Impacts on Classroom Climate at Preschool Follow-Up. For five measures of emotional support, four measures of classroom organization, four measures of instructional support, and one measure of literacy focus, the teachers in the pooled enhanced classrooms did not display any better or worse outcomes compared with the teachers in the regular Head Start classrooms.⁷⁷

Child-Level Impacts on Social-Emotional Outcomes at Preschool Follow-Up. For four of 12 (33.3 percent) assessments of social-emotional outcomes for the three-year-old cohort in preschool, the combined effect of the enhanced series had

75. Ibid., p. 26, Table 3.

76. Ibid.

77. Ibid., Table 6, pp. 33–34.

beneficial impacts that were statistically meaningful.⁷⁸ The enhanced services had small effects on social skills ($ES = 0.27, p = 0.05$), assertion behaviors ($ES = 0.25, p = 0.05$), self-control ($ES = 0.30, p = 0.05$), and closeness of student-teacher relationship ($ES = 0.22, p = 0.05$). While the enhancements had a small marginally significant impact on cooperation behaviors ($ES = 0.22, p = 0.10$), the pooled services had no effect on all four measures of behavior problems, degree of interpersonal skills, work-related skills, and conflict between students and teachers.

Child-Level Impacts on Three-Year-Old Pre-Academic Skills at Preschool Follow-up. For the teacher-reported pooled results, the enhancements failed to produce statistically meaningful differences among the three-year-old children, compared with their counterparts in Head Start as usual classrooms, on general knowledge, language and literacy,

and mathematical thinking.⁷⁹ While the main focus of the three enhancements was on children's social-emotional competence, the authors of the evaluation speculated that exposure to the enhancements would indirectly improve the pre-academic skills of three-year-olds.⁸⁰

In sum, the findings for the three-year-old cohort do not indicate that the pooled enhancements offer much of an improvement over traditional Head Start. However, the authors of the three-year-old cohort evaluation suggest some caution in drawing conclusions. First, "the conclusions that can be drawn from this analysis are limited because of the sample sizes, data sources, and measures available for the analysis."⁸¹ Second, "the pattern of impacts on 3-year-olds' social-emotional outcomes does not clearly align with the impacts on teacher practice and classroom climate in the classrooms serving these children."⁸²

78. Ibid., Table 7, p. 38.

79. Ibid., Table 9, p. 45.

80. Ibid., p. 44.

81. Ibid., p. ES-2.

82. Ibid.