2018-19 Evaluation of the Home Instruction for Parents of Preschool Youngsters (HIPPY) Program
Sydney Rowles, Ph.D. & Sarah Hinkley, Ph.D.

At-a-Glance

Home Instruction for Parents of Preschool Youngsters (HIPPY) is a home visiting program for parents of three-, four-, and five-year-olds that involves training parents to be their children’s first teachers. The program targets school readiness and parent involvement and is the largest of the 11 HIPPY programs in the state of Texas. HIPPY Dallas Independent School District (ISD) had a total budget of $2,408,737 for program year 2018-19, including $751,242 from the Dallas ISD general operating funds, $1,070,543 from Title I funds, and $586,952 from external sources.

Purpose of the Evaluation

The purpose of this evaluation was to evaluate and report outcomes of the HIPPY program. Specifically, this evaluation reports (1) student characteristic data for participants, (2) pre- and post-test results from a school readiness assessment, (3) kindergarten (K) readiness assessment outcomes, and (4) long-term reading and mathematics outcomes for former HIPPY participants.

What were the HIPPY program and student characteristics in 2018-19?

Methodology

HIPPY Dallas ISD program implementation information primarily came from an interview with the HIPPY Dallas ISD program director and data files from the Texas HIPPY office. The evaluators conducted frequency analyses to summarize participant characteristics.

Results

In 2018-19, HIPPY Dallas ISD served 1,253 children and 1,135 parents. Enrollment was highest in HIPPY Year 1 (45%). Most children (88%) and parents (89%) in the program were Hispanic. About half of the children were female (53%). Most children participated with a biological parent (96%). A majority of the parents were married (63%); completed a high school education, GED, or less (77%); and/or completed the program using the Spanish curriculum (81%).

A few more children participated in HIPPY Dallas ISD this year (N = 1,253) compared to 2017-18 (N = 1,231). The number of completers decreased slightly from 1,106 in 2017-18 (90%) to 1,060 in 2018-19 (85%).

What were the retention rates of HIPPY participants from 2017-18 to 2018-19?

Methodology

The evaluators used frequency analyses to determine the number of HIPPY children who remained enrolled in the program from 2017-18 to 2018-19.

Results

Of the 539 2017-18 HIPPY Year 1 children, 261 (48%) remained enrolled in 2018-19 as HIPPY Year 2 children. Of the 410 2017-18 HIPPY Year 2 children, 287 (70%) remained enrolled in 2018-19 as HIPPY Year 3 children.

Did HIPPY participants score at or above school-ready benchmarks?

Methodology

The evaluators used paired-sample t tests (statistical significance) and Cohen's d effect sizes (practical significance) to compare pre- and post-test scores on the English and Spanish versions of the Bracken School Readiness Assessment (BSRA-3).

Results

Composite test raw scores at all curriculum levels in both English and Spanish improved at a statistically significant level from pre-test to post-test (p < .001). All improvements were practically significant with Cohen’s d effect size values ranging from .70 (HIPPY Year 3 Spanish) to 1.34 (HIPPY Year 2 Spanish). Both English

---

1 Portions of this evaluation were adapted from Hinkley (2018).
2 As reported by the Early Learning Department on November 29, 2018. The grants office reported $1,372,131 in Title I funds for HIPPY as of March 28, 2019.
3 Texas HIPPY data files received July 03, 2019.
4 The HIPPY curriculum consists of three years: Year 1 (first year) through Year 3 (last year). Two children could not be classified by year and are not included in the percentage.
5 Forty-nine percent did not return, two percent repeated Year 1, and 0.4 percent completed Year 3.
6 Twenty-nine percent did not return and 0.5 percent repeated Year 2.
7 Cohen’s d values of 0.2, 0.5, and 0.8 are considered small, medium, and large effect sizes, respectively (Cohen, 1988).
8 Children with birthdates outside of the testable age range, or with missing pre- or post-tests, birthdates, or curriculum years were excluded from analyses.
and Spanish average standard scores for all curriculum years improved from pre-test to post-test (Figure 1). Average standard scores in both languages exceeded the school-ready threshold (standard scores > 85) by the end of the program year. Many HIPPY Year 1 (77% English; 54% Spanish), Year 2 (80% English; 83% Spanish), and Year 3 (82% English; 77% Spanish) children were classified as meeting age-appropriate standards by the end of the year.

Figure 1: 2018-19 HIPPY Dallas ISD Average BSRA-3 Standard Scores at Pre-Test and Post-Test

<table>
<thead>
<tr>
<th></th>
<th>English (n = 70)</th>
<th>Spanish (n = 302)</th>
<th>English (n = 54)</th>
<th>Spanish (n = 240)</th>
<th>English (n = 44)</th>
<th>Spanish (n = 190)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre</td>
<td>90.0</td>
<td>81.5 †</td>
<td>90.1</td>
<td>88.8</td>
<td>95.5</td>
<td>94.6</td>
</tr>
<tr>
<td>Post</td>
<td>97.6</td>
<td>87.4</td>
<td>97.7</td>
<td>98.1</td>
<td>97.1</td>
<td>96.1</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Texas HIPPY data files dated July 03, 2019. Note: BSRA-3 = Bracken School Readiness Assessment. † = This average standard score did not meet the age-appropriate cutoff (>85); Maximum possible score = 160 (Bracken, 2002). There was no normative sample of Spanish version test-takers. Refer to Palladino (2016) regarding limitations of the analysis and normative sample.

### Methodology

The evaluators examined Istation’s Indicators of Progress (ISIP) and the Texas Kindergarten Early Assessment (TX-KEA) to measure kindergarten readiness. ISIP is a computer-adaptive assessment that provides continuous monitoring of progress in reading skills. TX-KEA is a multidimensional assessment; however, the Emergent Literacy: Reading domain is considered the K readiness standard and is reported here. The evaluators used beginning-of-year (BOY) ISIP and BOY TX-KEA Emergent Literacy: Reading data to examine K readiness rates between former HIPPY students and matched non-HIPPY control groups. The matched control groups were created using propensity score matching to identify students who took the same exam but did not participate in the HIPPY program. The evaluators used chi-square analyses and Cohen’s w to test the differences in K readiness rates between the two groups for both assessments.

### Results

**ISIP**

A greater percentage of former HIPPY participants (66%; n = 351) were kindergarten ready as measured by BOY ISIP when compared to matched non-HIPPY participants (57%; n = 346). This difference was statistically and practically significant, \( \chi^2 (1) = 6.58, p = .013, \omega = .10 \). Although the HIPPY students were not statistically compared to the district (51%; n = 9,518), the HIPPY students also outperformed the district on BOY ISIP.

**TX-KEA**

A greater percentage of former HIPPY participants (86%; n = 360) were kindergarten ready as measured by BOY TX-KEA when compared to matched non-HIPPY participants (75%; n = 353). This difference was statistically and practically significant, \( \chi^2 (1) = 16.04, p < .01, \omega = .15 \). Although the HIPPY students were not statistically compared to the district (70%; n = 9,534), the HIPPY students also outperformed the district on BOY TX-KEA.

### Did the long-term academic outcomes of HIPPY participants differ from matched non-HIPPY participants?

**Methodology**

The evaluators used TerraNova/SUPERA (kindergarten through grade two) and State of Texas Assessments of Academic Readiness (STAAR; grade three) to compare normal curve equivalent (NCE: TerraNova/SUPERA) and mean scale scores (STAAR) between former HIPPY students and matched non-HIPPY control groups. The rates of students who were at or above the 40th percentile on participate in HIPPY but had the same probability of HIPPY enrollment. This methodology allows for group comparisons with reduced concern about systematic differences that may confound the effects of the treatment. For a detailed methodology on propensity score matching see Palladino (2016).

9 For the analysis, the English and Spanish administrations were combined into one file to allow for examination of overall student performance. Comparisons between the English and Spanish versions of ISIP are not recommended because the two versions are normed differently. Therefore, the percentage of students at or above Tier 1 (ISIP) or On Track (TX-KEA) were used; both indicate being kindergarten ready. BOY ISIP was dated November 5, 2018. BOY TX-KEA was dated November 1, 2018.

10 Non-HIPPY matched controls included students who took BOY ISIP and BOY TX-KEA in kindergarten who did not participate in HIPPY but had the same probability of HIPPY enrollment. This methodology allows for group comparisons with reduced concern about systematic differences that may confound the effects of the treatment. For a detailed methodology on propensity score matching see Palladino (2016).

11 Cohen’s \( \omega \) values of 0.1, 0.3, and 0.5 are considered small, medium, and large effect sizes, respectively (Cohen, 1988).

**Results**

**TerraNova/SUPERA**

HIPPY students outperformed non-HIPPY students on eight of nine TerraNova/SUPERA mathematics and reading exams (Figures 2 and 3).

For mathematics, the TerraNova K and grade-one passing rate differences between HIPPY and non-HIPPY students were statistically, but not practically, significant, $\chi^2 (1, n = 725) = 4.13$, $p = .04$, $w = .08$ for K and $\chi^2 (1, n = 749) = 6.10$, $p = .01$, $w = .09$ for grade one. The grade-two passing rate difference was not statistically significant.

Kindergarten and grade-one TerraNova mathematics NCE scores were higher for HIPPY students compared to non-HIPPY students. These differences were statistically and practically significant $t(723) = -2.03$, $p = .04$, $d = .15$ for grade one, and $t(747) = -2.92$, $p < .01$, $d = .21$ for grade two. The difference in grade-two NCE scores was not statistically significant.

For reading, the difference in grade-one passing rates between HIPPY and non-HIPPY students on SUPERA was statistically significant, $\chi^2 (1, n = 594) = 4.03$, $p < .05$, $w = .08$. No other passing rate differences on TerraNova or SUPERA were statistically significant.

Kindergarten and grade one SUPERA reading NCE scores were higher for HIPPY students compared to non-HIPPY students. These differences were statistically and practically significant $t(587) = -3.06$, $p < .01$, $d = .25$ for K and $t(592) = -2.32$, $p = .02$, $d = .19$ for grade one. Differences in all other TerraNova and SUPERA NCE scores were not statistically significant.

Although the HIPPY students were not statistically compared to the district, the HIPPY students also outperformed the district on all exams.
STAAR

Higher percentages of former HIPPY than non-HIPPY students scored at the Approaches+ level on STAAR mathematics and reading (Table 1), but none of these differences were statistically significant.

Similarly, higher percentages of former HIPPY than non-HIPPY students scored at the Meets+ level on STAAR reading (Figure 4). However, the reverse was true for STAAR mathematics. Neither of these differences were statistically significant.

Although it was not appropriate to statistically compare HIPPY students to the district as a whole because of differing sample sizes and possible confounding factors, the percentage of HIPPY scores at Approaches+ and Meets+ exceeded the district for both STAAR exams.

Table 1: 2018-19 STAAR Mathematics and Reading Rates of Students at Approaches+ and Meets+ for Former HIPPY Participants, Matched Non-HIPPY Comparisons, and the District

<table>
<thead>
<tr>
<th></th>
<th>Approaches+</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>STAAR Mathematics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIPPY</td>
<td>216</td>
<td>76.9</td>
</tr>
<tr>
<td>Non-HIPPY</td>
<td>215</td>
<td>73.5</td>
</tr>
<tr>
<td>District</td>
<td>10,955</td>
<td>76.6</td>
</tr>
<tr>
<td>STAAR Reading</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIPPY</td>
<td>216</td>
<td>75.5</td>
</tr>
<tr>
<td>Non-HIPPY</td>
<td>214</td>
<td>74.3</td>
</tr>
<tr>
<td>District</td>
<td>10,960</td>
<td>71.9</td>
</tr>
</tbody>
</table>


Note: STAAR = State of Texas Assessments of Academic Readiness. Approaches+ = Approaches Grade Level or Above. Meets+ = Meets Grade Level or Above. Data include first and second administrations of all versions of STAAR. No differences between HIPPY and non-HIPPY control groups were statistically significant.

No differences between HIPPY and non-HIPPY groups on STAAR average scale scores were statistically significant.

Recommendations

• **Continue to seek areas for improvements in effectiveness.** The program has enjoyed pre/post gains in developmental skills every year it has been evaluated. HIPPY leadership should continue to nurture its staff and review processes to maintain these strong results.

• **Include HIPPY in the evaluation of other major Dallas ISD early childhood programs.** Many children in the Dallas ISD attend more than one pre-kindergarten program. Thus, it has become more difficult to evaluate the efficacy of any one early childhood program exclusively, particularly for long-term outcomes. Evaluation and program managers should consider program participation and dosage as a variable in future evaluations of early childhood education. Data structures and systems should be carefully reviewed and enhanced to support this goal.