Parents as Teachers Home Visiting Intervention: A Quasi-Experimental Evaluation of Academic Outcomes, School Behavior, and Parenting Skills

August 2018

LeCroy & Milligan
ASSOCIATES, INC.

Parents as Teachers Home Visiting Program: A Quasi-Experimental Evaluation of Academic Outcomes, School Behavior, and Parenting Skills

August 2018

Submitted to:
Parents as Teachers, Inc.
2228 Ball Drive
St. Louis, MO 63146
Ph: (866) 728-4968
Fax: (314) 432-8963
ParentsAsTeachers.org

Submitted by:
LeCroy & Milligan Associates, Inc.
2002 N. Forbes Blvd. Suite 108
Tucson, AZ 85745
Ph: (520) 326-5154
Fax: (520) 326-5155
www.lecroymilligan.com

Acknowledgments:

The evaluation team for this project would like to thank Allison Kemner, Vice President for Research and Quality, PATNC, and program administrators for PAT at Sunnyside Unified School District, Joan Katz and Adriana Huerta, for their extensive efforts with coordination, oversight, and guidance of this study. The evaluation team includes Greg Goodman, MS, Michele Schmidt, MPA, and Michel Lahti, PhD.

About LeCroy & Milligan Associates:

Founded in 1991, LeCroy & Milligan Associates, Inc. is a consulting firm specializing in social services and education program evaluation and training that is comprehensive, research-driven and useful. Our goal is to provide effective program evaluation and training that enables stakeholders to document outcomes, provide accountability, and engage in continuous program improvement. With central offices located in Tucson, Arizona, LeCroy & Milligan Associates has worked at the local, state and national level with a broad spectrum of social services, criminal justice, education and behavioral health programs.

Suggested Citation:

LeCroy & Milligan Associates, Inc. (2018). *Parents as Teachers Home Visiting Intervention: A Quasi-Experimental Evaluation of Academic Outcomes, School Behavior, and Parenting Skills*. Tucson, AZ.

Table of Contents

List of Exhibits

Executive Summary

The aim of this study was to estimate the program effect of the Parents as Teachers (PAT) home visiting intervention on Sunnyside Unified School District (SUSD) students and their parents who participated in the program (i.e., PAT group), relative to a comparison group of SUSD students who did not participate in the PAT program (non-PAT group). Using a matched comparison group quasi-experimental design, this study assessed three years of academic and school outcomes in the areas of:

- Reading and math standardized assessment scores;
- Achievement of English language proficiency of English Language Learner (ELL) students; and
- School indicators of absentee rates and number of suspensions.

This study also examined change in parenting skills of parents who participated in the PAT program from baseline to posttest assessment using the Keys to Interactive Parenting Skills (KIPS) and the Protective Factors Survey (PFS).

Findings on Student Achievement in ELA/Reading: PAT students demonstrated significantly higher AzMERIT ELA/reading assessment outcome scores from 2015 to 2017 than the matched non-PAT comparison group ($p < .05$). PAT students also performed significantly better than the matched non-PAT comparison group on RAPS 360 Phonics (in 2015, 2016, and 2017) and Comprehension (in 2017) subscales ($p < .01$). However, total RAPS 360 fluency scale scores and other subscale scores did not show a significant difference over time between the two groups.

Findings on Student Achievement in Math: PAT students demonstrated a significantly higher math gain score from 2015 to 2017 on the AzMERIT math scale scores than the matched non-PAT comparison group ($p < .001$).

Findings on AZELLA Reading, Writing and Total Scores: PAT ELL students significantly outperformed non-PAT ELL students in Reading, Writing and Total in 2015; Writing and Total in 2016; and Writing in 2017. ($p < .05$).

Findings on School Indicators of Absence Rates and Days Suspended: PAT participants had significantly lower average rates of absenteeism than non-PAT comparison students in 2014, 2015, 2016, and 2017 ($p < .000$). PAT participants had significantly fewer out-of-school suspension days than non-PAT comparison students in 2015. ($p < .05$).

Findings on PAT Parent Assessments: PAT parents showed significant improvement in parenting quality over time, based total mean scores from pretest to posttest as measured by the Keys to Interactive Parenting Scale (KIPS). PAT parents also showed significant improvements over time in three of the four the Protective Factors Survey (PFS) domains of family functioning, social support, and concrete support from pretest to posttest. For significant assessments, the posttest mean of parents after program implementation was statistically significantly higher than their pretest means at the start of the program ($p < .01$).

A summary of key findings is shown in the infographic on the next page of this report. This study shows promising results that participation in the PAT program in early childhood may lead to better academic outcomes in the future. Likewise, assessments of PAT parents showed a statistically significant improvement in parenting quality after program participation, which likely supported the positive academic gains realized by their children. Giving the positive findings of this study, further investigation of how participation in the PAT program impacts child and parent outcomes is warranted. Additional co-variates could include measuring program fidelity to the PAT National Center (PATNC) model and dosage of programming received (e.g., frequency of home visits, duration of program participation). As PATNC continues to collaborate with SUSD, additional data available in the coming years will allow for more rigorous analysis. Once a larger cohort of PAT participants advances into higher grades and completes academic assessments, there will be a better opportunity to evaluate relationships and differences between school outcomes and variables predictive of PAT participation.

Sunnyside Unified School District Parents as Teachers Program Effect on Student and Parenting Outcomes: Key Study Findings

Outcome Area	Research Question		Result
ELA/Reading Achievement	Is there a difference in reading outcomes between PAT and non-PAT students as measured by the AzMERIT ELA assessment?		PAT students demonstrated significantly higher AzMERIT ELA/reading assessment outcome scores from 2015 to 2017 than the matched non-PAT comparison group.
	Is there a difference in reading outcomes in fluency between PAT and non-PAT students as measured by RAPS 360?		PAT students performed significantly better than the matched non-PAT comparison group on RAPS 360 Phonics (in 2015, 2016, and 2017) and Comprehension (in 2017) subscales.
Math	Is there a difference in math outcomes between PAT and non-PAT comparison students as measured by the state standardized math test?		PAT students produced a significantly higher gain score from 2015 to 2017 on AzMERIT math scale scores than the matched non-PAT comparison group.
English Language Proficiency	Is there a difference in reading outcomes between PAT and non-PAT students as measured by the AZELLA?		PAT ELL students significantly outperformed non-PAT ELL students on the AZELLA in Reading, Writing and Total in 2015; Writing and Total in 2016; and Writing in 2017.
Absenteeism	Is there a difference between PAT and non-PAT students on child school indicators of absence rates in the last three school years?		PAT participants had significantly lower average rates of absenteeism than non-PAT comparison students in 2014, 2015, 2016, and 2017.
Suspensions	Is there a difference between PAT and non-PAT students number of suspensions (in-school and out-of-school) in the last three school years?		PAT participants had significantly fewer out-of-school suspension days than non-PAT comparison students in 2015.
Parenting Quality	To what extent does the PAT intervention improve parenting skills from pre to post intervention?		PAT parents demonstrated significantly improved parenting quality from their initial KIPS assessment to a subsequent assessment performed over the course of the program.
Protective Factors	To what extent does the PAT intervention improve four protective factor domains from pre to post intervention?		PAT parents showed significant improvements over time in three of the four protective factor domains: family functioning, social support, and concrete support.

Introduction

LeCroy & Milligan Associates was commissioned by Parents as Teachers National Center, Inc. to conduct an outcome study that examines the effect of the Parents as Teachers (PAT) home visitation program intervention on parents and children in Sunnyside Unified School District (SUSD) in Pima County, Arizona, compared with matched-comparison group of similar children in SUSD who did not participate in the PAT program. The PAT program has been in operation in the SUSD since 1996 and serves more than 400 families annually, providing information, support, and encouragement for parents to help their children develop optimally during the crucial early years before school. Affiliates need to demonstrate annually that they are meeting PAT Essential Requirements by submitting data to the PAT National Center. Every 5 years, affiliates continue to go through this Quality Endorsement and Improvement Process. In 2017, the SUSD PAT program achieved Blue Ribbon Affiliate Status by meeting 85 of 100 quality standards. The SUSD PAT has consistently maintained this level of quality since 2016.

Overview of the Parents as Teachers Program Model

PAT is an innovative, home-based intervention that provides childhood family support and parent education beginning with pregnancy and extending until entry into kindergarten (PAT, 2018). Certified Parent Educators (PE) are at the heart of the PAT model. PEs visit parents and children at home for approximately one hour every week, every other week, or monthly, depending upon the specific needs of the family. The Parents as Teachers National Center (PATNC) has developed the comprehensive research- and evidence-based *Foundational Curriculum* (2017), covering prenatal to three years, and *Foundational 2 Curriculum* (2014), covering three years through Kindergarten. This curriculum is implemented as part of the PAT model during each visit with the following goals:

(1) Increasing parent knowledge of early childhood development (e.g., parents are taught to identify and encourage the development of age appropriate milestones in language, cognitive, social-emotional, and motor skills);

(2) Improving parenting practices (e.g., parents practice positive parent-child interactions that are nurturing, responsive, and supportive of the child's learning and development);

(3) Providing early detection of developmental delays and/or health issues (e.g. PAT children are given a health screening that evaluates health status, safety, vision, and hearing, in addition, a developmental screening is completed that assesses language, cognitive, social-emotional, and motor development. Child development is then monitored during each subsequent visit);

(4) Preventing child abuse and neglect (e.g., strengthening families by developing protective factors, such as parental resilience, social connectedness, and concrete support); and

(5) Increasing children's school readiness and school success (e.g., teaching parents to promote healthy child brain development through daily play and hands-on experiences).

The curriculum focuses on three central developmental topics: (1) parent-child interactions (i.e., increasing positive parenting behaviors and promoting child development through parent-child activities); (2) development centered parenting (i.e., the connection between child development and parenting including a focus on attachment, discipline, health, nutrition, safety, sleep, transitions/routines, healthy births); and (3) family well-being (i.e., family strengths, capabilities, skills, fostering protective factors). In addition to these structured home visits, PAT provides families with monthly (or more frequent) group gatherings including family activities, presentations, community events, parent cafes, and group meetings. These venues provide families with social support and the opportunity to obtain information and share experiences with their peers. Further, PAT provides annual health, hearing, vision, and developmental screenings, starting within 90 days of enrollment. For the SUSD PAT program, the Parent Educators are bi-lingual and conduct home visits in the family's preferred language. Additionally, curriculum and materials are available in both English and Spanish languages. The *PAT Foundational Training Guide* (PATNC, 2015) includes several modules on cultural competency, encouraging parent educators to understand the cultures of the families served in order to provide culturally appropriate parent education.

Study Purpose

The aim of this study was to estimate the program effect of the PAT home visiting intervention on SUSD students and their parents who participated in the program (i.e., PAT group), relative to a comparison group of SUSD students who did not participate in the PAT program (non-PAT group). Using a matched comparison group quasi-experimental design, this study assessed three years of academic and school outcomes in the areas of:

- Reading and math standardized assessment scores;
- Achievement of English language proficiency of English Language Learner (ELL) students; and
- School indicators of absentee rates and number of suspensions.

This study also examined change in parenting skills of parents who participated in the PAT program from baseline to posttest assessment (within group change) on two parenting measures: Keys to Interactive Parenting Scale (KIPS) and the Protective Factors Survey (PFS).

Methodology

Study Design

The evaluation of PAT, as implemented in SUSD, uses a matched comparison group quasi-experimental design. Analysis of student outcomes is guided by the U.S. Department of Education *What Works Clearinghouse* (WWC) recommendations on how best to conduct student achievement research (WWC, 2017). These recommendations are applied as part of a between-groups design where the treatment group (families who received PAT programming) and a comparison group (families who did not receive PAT programming) on student achievement and reading assessments are evaluated. Students in grades pre-K through 12th who participated in the PAT program were matched to comparison students in SUSD, based on key demographic variables: gender, ethnicity, special education students (SPED), English Language Learner (ELL) students, and receipt of free or reduced-price lunch. Propensity Score matching was utilized to match one treatment case for every five comparison cases (1:5) based on each case's propensity score (Randolph, Falbe, Manuel, & Balloun, 2014). The evaluation presents results comparing outcomes of PAT and non-PAT students for three school fiscal years (SY) 2014-2015 (Year 1), 2015-2016 (Year 2), and 2016-2017 (Year 3).

Research Questions

There is limited research establishing if and how PAT participation effects specific forms of academic achievement, such as reading and math abilities, and no past research on the effect that PAT has on school absenteeism or suspensions. This study fills these gaps by answering the following research questions:

(1) To what extent does the PAT intervention result in statistically significant increases in student reading and math assessment scores compared to a comparison group of non-PAT students?

(2) To what extent do English Language Learner (ELL) students that receive the PAT intervention achieve English language proficiency compared to a comparison group of ELL non-PAT participants?

(3) Is there a difference between PAT and non-PAT students on child school indicators of absentee rates and number of suspensions (in-school and out-of-school) in the last three school years?

(4) To what extent does the PAT intervention improve parenting skills from baseline to posttest assessment (within group change) on two parenting measures?

Based on past research, it was hypothesized that the PAT group would outperform the non-PAT group on each of the outcomes and that parenting post-test scores would be significantly higher than pre-test scores.

Study Measures

Student Outcomes

This study examined the impact of the PAT SUSD intervention on PAT and non-PAT SUSD students in the areas of English Language Arts (ELA)/reading and math achievement. The student achievement outcomes, measurement, and tools utilized for this study are summarized in Exhibit 1.

Exhibit 1. Student Achievement Outcomes and Measurement Tools

Outcomes Measured	Measure	Measurement Tool	Tool Description
English Language Arts (ELA)/Reading	Reading Score	AzMERIT	AzMERIT is an annual statewide test that was administered to Arizona students every spring for SY2015, 2016, and 2017
	Phonics, Comprehension, and Fluency Scores	Reading Analysis and Prescription System (RAPS) 360	RAPS 360 is a computerized reading assessment program administered to students to identify strengths and weaknesses in Phonics, Comprehension, and Fluency.
English Language Proficiency	Reading Score, Writing Score, Total Score	Arizona English Language Learner Assessment (AZELLA)	Arizona English Language Learner Assessment (AZELLA) is a standards-based assessment with state and federal requirements to measure students' English language proficiency
Math	Math Score	AzMERIT	AzMERIT is an annual statewide test that was administered to Arizona students every spring for SY2015, 2016, and 2017
School Outcomes	Absence Rates In and Out-of-School Suspensions	SUSD Administrative Data	Administrative data submitted by SUSD and matched to PAT and non-PAT study groups.

Arizona Measurement of Educational Readiness to Inform Teaching (AzMERIT)

The Arizona Measurement of Educational Readiness to Inform Teaching (AzMERIT; Arizona Department of Education [ADE] & American Institutes for Research [AIR], 2017) is an annual statewide student achievement test administered every Spring beginning in SY2015. This test assesses ELA/reading achievement and math achievement for students in grades 3-11. For the current study, the SY2015 AzMERIT ELA and math assessment scores were used as the initial measures for comparing PAT and non-PAT students in the analytic sample. AzMERIT data collected in SY2016 and SY2017 were used as the comparison data to examine change over time. The AzMERIT is an online assessment taken in two or three discrete test sessions. The instrument's items include a variety of selected responses, machine scored constructed responses (i.e., graphic response, natural language, equation response, hot text, table input items), and essay responses. The Cronbach's alpha for the ELA and math AzMERIT is

uniformly in the 0.90 range, which is consistent with most other similar length achievement tests (ADE & AIR, 2017; Cronbach, 1951; Nunnaly, 1978[1]).

The Reading Analysis and Prescription System (RAPS) 360

The Reading Analysis and Prescription System (RAPS 360; Mindplay & Methods and Solutions, Inc., 2015) is a computerized reading assessment program administered to students to identify strengths and weaknesses in various areas of reading. The RAPS 360 assesses eight domains of reading including: Comprehension, Phoneme Segmentation, Listening Vocabulary, Phonics-Decoding/Encoding, Visual Scanning Efficiency, Natural Fluency, Expected Fluency, and the Pause-Assisted Fluency and takes between 20-45 minutes to complete. The current study evaluated RAPS 360 results in the three specific areas of Comprehension, Phonics, and Fluency. The Fall SY2014 benchmark assessments in these three areas were used as baseline measures for comparing PAT and non-PAT students in the analytic sample.

Arizona English Language Learner Assessment (AZELLA)

The Arizona English Language Learner Assessment (AZELLA; ADE & Harcourt Assessment, Inc., 2007) measures students' English language proficiency; students who score *Proficient* on the test are deemed to have sufficient English proficiency to be placed in mainstream, English speaking classrooms (ADE, 2014). The AZELLA assesses listening, speaking, reading, and writing ability; the current study evaluated the reading and writing subscales along with AZELLA total scores, with SY2014 as the baseline measure for comparing PAT and non-PAT students in the analytic sample. Past research indicates that the Cronbach's alpha for the total AZELLA scores was 0.87 and the average machine Cronbach's alpha across grades was 0.83 (ADE, 2016). AZELLA scores showed strong internal consistency with Cronbach alpha scores ranging from .93 to .97.

School Attendance and Suspensions

SUSD administrative data on number of absences, in-school suspensions, and out-of-school suspensions were analyzed to compare school attendance and behavior for PAT participants versus non-PAT participants for SY2015, 2016, and 2017.

Parent Outcomes

Keys to Interactive Parenting Scales (KIPS)

The Keys to Interactive Parenting Scales (KIPS) is a validated, structured observational assessment that examines caregiver-child interactions during play (Comfort & Gordon, 2006; Comfort, Gordon, & Naples, 2011; Comfort et al., 2010; Comfort, Gordon & Unger, 2006). The KIPS was completed annually by PAT staff and was used to determine the extent to which the PAT intervention improved skills in parent-child relationships, learning, and confidence as parents. The data were analyzed as pretest/posttest paired scores. The KIPS is an observational

[1]Cronbach (1951) and Nunnaly (1978) report that a Cronbach alpha score of .70 or higher demonstrates strong internal consistency or average correlation of items in a survey instrument.

measure that assesses a caregiver's interaction with a child over a 20-minute time period. Scores were obtained by PAT staff who were trained in the use of the KIPS, however, there was no assessment of inter rater reliability of this measure for this part of the study. KIPS developers report results of high interrater reliability (*r=.88),* and good coefficient alphas of 0.89. The KIPS has shown significant positive correlations with Nursing Child Assessment Teaching Scale (NCATS) and the Home Observation for Measurement of the Environment (HOME) subscale scores (Comfort & Gordon, 2006).

Protective Factors Survey (PFS)

The Protective Factors Survey (PFS; Counts, Buffington, Chang-Rios, Rasmussen, & Preacher, 2010) is a self-administered pre-post evaluation tool designed for use with caregivers receiving child abuse prevention services and to provide feedback to programs for continuous improvement and evaluation purposes. Results are designed to identify changes in protective factors and measure individual family protective outcomes. The PFS has 20 items that assess the domains of family functioning, social support (i.e., from family and friends), concrete support (i.e., access to tangible goods and services), nurturing and attachment (i.e., positive interactions between parent and child), and knowledge of parenting and child development. For this current study, subscale reliability ranged from 0.74 to 0.85. These results are consistent with Counts et al. (2010) who reported subscale reliability ranging from 0.76 to 0.89, internal consistency = 0.93, and test–retest reliability ranging from 0.63 to 0.88.

Study Samples and Demographics

Exhibit 2 shows the sample sizes for each study group by outcome area and measure examined in this study.

Exhibit 2. Summary of Study Samples by Outcome Area and Measure

Outcome Area	Measure	PAT Treatment Group N	Non-PAT Comparison Group N
Student Outcomes	AzMERIT ELA/Reading and Math	625	3,125
	RAPS 360 School Attendance and Suspensions	983	4,915
	AZELLA	60	229
Parent Outcomes	KIPS	182	-
	PFS	250	-

SUSD Student Outcome Samples

De-identified administrative data from SUSD was utilized to generate the PAT treatment and non-PAT comparison groups of students with comparable student outcomes. It is important to note that each study sample represents distinct data sets for analysis and are not compared with each other, due to a limited overlap of outcome data and intervention data.

AzMERIT Outcomes Study Sample

A total of 625 SUSD students who participated in the PAT program had state standardized test data in ELA/reading and math (i.e., AzMERIT) from spring 2015, 2016, and 2017. Propensity Score matching was performed with the larger SUSD sample of non-PAT students to create a well-matched comparison group of 3,125 SUSD students with state standardized test data who did not participate in PAT. The age of students ranged from 6 years to 18 years, in grades 3 through 11. Exhibit 3 shows the demographic characteristics for the PAT and non-PAT comparison groups in the AzMERIT outcomes study sample. Demonstrating the success of the Propensity Score Matching technique used, Chi square tests showed no significant difference (p=1.00) between these two groups and their demographic characteristics for this study sample.

Exhibit 3. Demographic Characteristics of PAT and Non-PAT Students in the AzMERIT Outcomes Study Sample and Chi Square Baseline Equivalence Test Results

Demographic Characteristic		AzMERIT Outcomes Study Sample (n = 3,750) PAT (n = 625)	Non-PAT (n = 3,125)	x^2	p
Gender	Female	51.7%	51.7%	.000	1.00
	Male	48.3%	48.3%		
Race/ Ethnicity	Hispanic	97.8%	97.8%	.000	1.00
	White	0.6%	0.6%		
	African American	0.2%	0.2%		
	Native American	1.0%	1.0%		
	Asian/Pacific Islander	0.4%	0.4%		
Other	Special Education Student	12.5%	12.5%	.000	1.00
	English Language Learner	13.6%	13.6%	.000	1.00
	Free/Reduced Lunch Eligible	86.1%	86.1%	.000	1.00

RAPS 360 and School Outcomes Study Sample

In addition to the AzMERIT outcomes study sample, SUSD data was used to generate a second dataset of PAT and non-PAT comparison students who had comparable reading assessment scores from the RAPS 360 and absenteeism and suspension data from 2014–2017. This dataset included 983 SUSD pre-school and elementary school students who participated in PAT and a well-matched comparison group of 4,915 SUSD students who did not participate in PAT. Exhibit 4 shows the demographic comparison for PAT and non-PAT comparison students in this study sample. Demonstrating the success of the Propensity Score Matching technique used, Chi square tests showed no significant difference (p=1.00) between these two groups and their demographic characteristics.

Exhibit 4. Demographic Characteristics of PAT and Non-PAT Students for the RAPS 360 and School Outcomes Study Sample and Chi Square Baseline Equivalence Test Results

Demographic Characteristic		RAPS 360 and School Outcomes Study Sample (n = 5,898)			
		PAT (n = 983)	Non-PAT (n = 4,915)	x^2	p
Gender	Female	48.6%	48.6%	.000	1.00
	Male	51.4%	51.4%		
Race/ Ethnicity	Hispanic	96.3%	96.3%	.000	1.00
	White	1.3%	1.3%		
	African American	0.2%	0.2%		
	Native American	2.0%	2.0%		
	Asian/Pacific Islander	0.1%	0.1%		
Other	Special Education Student	8.2%	8.2%	.000	1.00
	English Language Learner	11.8%	11.8%	.000	1.00
	Free/Reduced Lunch Eligible	75.3%	75.3%	.000	1.00

AZELLA Study Sample

Exhibit 5 shows the demographic comparison for PAT and non-PAT comparison students in the AZELLA study sample of students that had data for SY 2015-2017. Chi square tests showed no significant difference between these two groups and their demographic characteristics.

Exhibit 5. Demographic Characteristics of PAT and Non-PAT Students for AZELLA Sample and Chi Square Baseline Equivalence Test Results

Demographic Characteristic		AZELLA Study Sample (n = 289)			
		PAT (n = 60)	Non-PAT (n = 229)	x^2	p
Gender	Female	47%	36%	2.183	.14
	Male	53%	64%		
Race/ Ethnicity	Hispanic	100%	100%	.000	1.00
Other	Special Education Student	18%	25%	1.000	.32
	Free/Reduced Lunch Eligible	95%	98%	2.129	.15

PAT Parent/Guardian Participants

In addition to the datasets to examine student-level outcomes, this study examined change in parenting quality and skills over time with a sample of 182 PAT parent/guardians with matched pre and post KIPS scores and a sample of 250 PAT parent/guardians with matched pre and post PFS subscale scores.

Data Analysis

Analysis on student outcomes was guided by What Works Clearinghouse (WWC) recommendations (U.S. Department of Education, WWC, 2017) and applied as part of a quasi-experimental, between groups design where treatment (PAT) and comparison (non-PAT) groups on student reading and math achievement, as well as school attendance and behavior, were evaluated for change over time. The WWC recommends establishing a Baseline Equivalence estimate and suggests guidelines for implementing statistical adjustments (See Ho, Imai, King, and Stuart, 2007). The WWC (2017) also recommends assessment of the practical importance of an intervention's effect by translating effect sizes into "improvement index" values. The improvement index represents the difference between the percentile rank consistent with the mean value of the outcome for the intervention group and the percentile rank consistent with the mean value of the outcome for the comparison group distribution. The improvement index represents the expected change in percentile rank for an average control group student if the student had received the intervention. Finally, The WWC (2017) has adopted the use of effect size estimates known as Hedges' g or Cohen's d. It is defined as the difference between the outcome (posttest) means for the intervention group and comparison group, divided by the unadjusted pooled standard deviation of the outcome measure. An effect size of 0.2 is considered small, 0.5 moderate, and 0.8 large (Cohen, 1988).

Propensity Score Matching

Because random assignment was not conducted, propensity score analysis was used to control for possible selection bias and ensure that participants in the PAT group were statistically indistinguishable across observed covariates from participants in the non-PAT group. A propensity score is the predicted probability of receiving treatment given observed covariates (Rosenbaum & Rubin, 1983). To demonstrate that the propensity score technique was successful in matching the two groups on key demographic variables of gender, ethnicity, special education, ELL and free/reduced lunch, a non-parametric Mann-Whitney U analysis was performed. This test identifies differences in mean ranks as well as the observed significance level (p -value) between the treatment and comparison groups for samples one (1) and two (2). The test indicated that for all key demographic variables, PAT participants and non-PAT comparisons were well-matched, sharing the same Mean Rank (1875.50), U = 976562.50, and p = 1.00. The same method was used for sample two (2) and again, the two groups were well-matched, sharing the same Mean Rank (2949.50), U = 2415722.50, and p = 1.00. Additionally, cross tabulation and Chi Square analysis was performed for each study sample by demographic area, to ensure that no significant differences were observed between the two study groups. These findings are important because the more similar the two groups are at baseline, the more observed differences between the two groups could be attributed to the intervention.

Statistical Tests

To examine the effect of PAT on student ELA achievement and reading assessment, Analysis of Covariance (ANCOVA) was used with the relevant ELA/reading score as the dependent variable and the PAT/non-PAT study groups as the independent variable. The analysis controlled for the baseline achievement measure in the same domain as the outcome.

To examine the effect of the PAT intervention on math achievement, gain score analysis was utilized because baseline mean differences between PAT and non-PAT comparison students was greater than 25% of the pooled standard deviation. The PAT/non-PAT study groups were evaluated for significance differences and effect size. Additionally, a non-parametric Mann-Whitney U analysis was also conducted on grade-level scores of the RAPS 360 Phonics, Comprehension, Phonemic Awareness, and Listening/Vocabulary Level to test for differences in mean ranks between the PAT/non-PAT study groups at the end of SY2014-2017.

A one-way multivariate analysis of variance (MANOVA) was conducted to test the hypothesis that there would be one or more mean differences between PAT students and non-PAT comparison students in AZELLA reading, writing, and total scores in SY2015, 2016, and 2017. Prior to conducting the MANOVA, a series of Pearson correlations were performed between all dependent variables to test the MANOVA assumption that the dependent variables would be correlated with each other in the moderate range (i.e., 0.20 - 0.60; Meyers, Gampst & Guarino, 2006).

Independent samples t-tests were conducted to examine the effect of the PAT program on absence rates and school suspensions in SY2014-2017, as school indicators of attendance and behavior. Mean absence rates and in-school and out-of-school suspension days were compared by the PAT/non-PAT study groups as the independent variable. Cohen's *d* was calculated to determine the effect size.

Prior to conducting the analysis on both KIPS and the PFS, the assumption of normally distributed difference scores was examined. The assumption was confirmed for both assessments, as the skew and kurtosis levels were less than the maximum allowable values for a t-test (i.e., skew < |2.0| and kurtosis < |9.0|; Posten, 1984). As per the developers of KIPS, the total mean KIPS score was calculated and interpreted in the following way: a mean score of ≤ 2.9 is a low score, indicating low quality parenting; a mean score of 3.0 - 3.9 is a medium score, indicating medium quality parenting; and a mean score of ≥ 4.0 is a high score, indicating high quality parenting. A dependent samples t-test was performed with paired pretest and posttest total mean KIPS scores to examine change over time. The PFS was also analysed in accordance with the developer's guidelines. Subscale scores were calculated for the family functioning, social support, concrete support, and nurturing and attachment domains and a dependent samples t-test was performed with paired pretest and posttest subscale scores to examine change over time.

Results

English Language Arts/Reading Achievement - AzMERIT

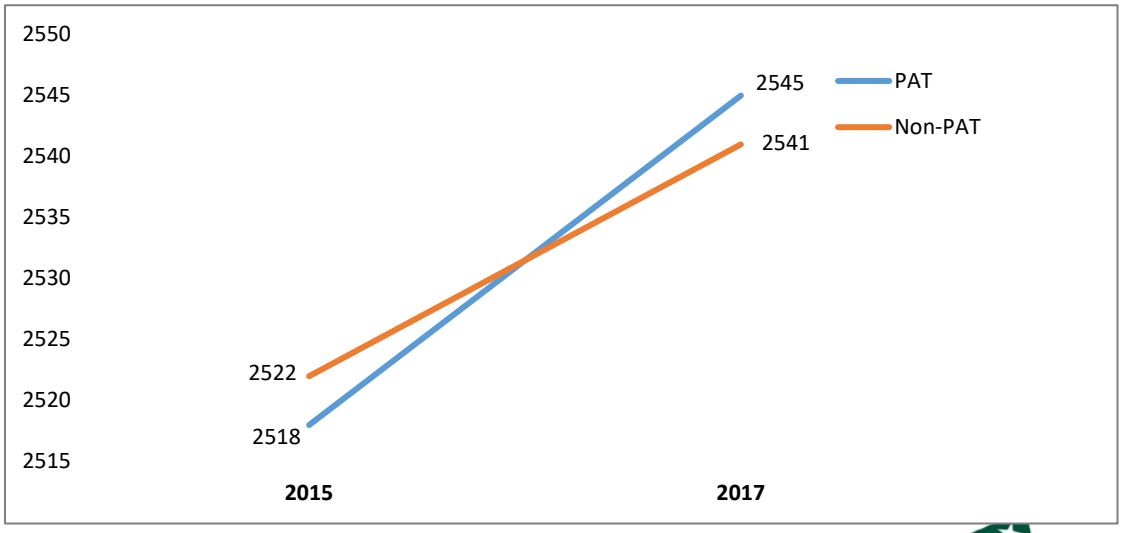
Is there a difference in reading outcomes between PAT and non-PAT students as measured by the AzMERIT ELA assessment? The results showed a difference in AzMERIT ELA/reading achievement outcome scores, with PAT students performing better than the matched non-PAT comparison group (see Exhibits 6 and 7). ANCOVA was performed because mean baseline differences were less than 25% of the PSD between the PAT group (n = 382) and the comparison group (n = 1175). ANCOVA showed a statistically significant difference in the adjusted 2017 mean scores between PAT participants (M = 2544.64; SD = 32.22) and non-PAT comparisons (M = 2541.64; SD = 30.44) after controlling for the 2015 score covariate. The fixed factor 'group' was significant, $F(1, 1554) = 6.002$, $p = .01$ as was the covariate 2015 score, $F(1, 1556) = 1883.41$, $p < 0.001$. The homogeneity of variance assumption between the two groups was confirmed with a Levene's test of p=.591. Hedges' g effect size calculation = 0.10, which is a small effect size based on Ellis' (2010) guidelines. The resulting improvement index was +4, corresponding to moving performance for the average student from the 50th to the 54th percentile of the comparison group distribution if the student had received the intervention.

Exhibit 6. AzMERIT ELA Mean Scores for PAT and Non-PAT Comparison Group Students, 2015 to 2017

Group	N	2015 AzMERIT ELA Score *Mean (SD)*	2017 AzMERIT ELA Score *Mean (SD)*	*F*	*p*	*ES*	Improvement Index
PAT	382	2518.12 (33.22)	2544.64 (32.22)	6.002	**.01**	.10	+4
Non-PAT	1175	2522.11 (32.78)	2541.64 (30.44)				

Note: ES = The adjusted posttest means and unadjusted standard deviations were used to compute Hedges' g effect size (.10), p = .014. Improvement Index = the expected change in percentile rank for an average comparison group student if the student had received the intervention. It is measured as the percentile difference between the intervention group mean and the comparison group mean using the comparison group distribution. P-values are deemed significant if ≤ .05. Significant p-values are shown in bold font.

Exhibit 7. Line Graph of AzMERIT ELA Mean Scores for PAT and Non-PAT Comparison Group Students, 2015 to 2017



Math Achievement - AzMERIT

Is there a difference in math outcomes between PAT and non-PAT comparison students as measured by the state standardized math test? Results showed a difference in AzMERIT math gain scores, with PAT students performing better than non-PAT comparison students (see Exhibit 8). The PAT participants demonstrated a statistically significant larger gain in mean math scale scores than the comparison students from 2015 to 2017. Gain Score Analysis was used because 2015 mean scale score difference was greater than 25% of the PSD (baseline equivalence on showed that the difference in 2015 mean scale scores by study group was 32% of the PSD). The PAT participant group (n = 364) had a 2015 mean scale score of M = 3567.07 (SD = 50.15) and a 2017 mean scale score of M = 3623.97 (SD = 46.89); by contrast, the non-PAT comparison group (n = 1193) demonstrated a larger 2015 mean scale score (M = 3584.75; SD = 57.28) and 2017 mean scale score (M = 3632.93; SD = 50.82).

An independent samples t-test was performed on gain scores calculated from 2015 to 2017 mean scale scores for each study group. The independent samples t-test on gain scores demonstrated a statistically significant effect, $t(1555) = 4.67$, $p = 0.000$. Thus, PAT participation was associated with a statistically significantly larger gain in mean math scale scores from 2015 to 2017 compared to non-PAT comparison students. Cohen's d (effect size) was estimated at 0.28 (a small to moderate effect size [0.2 is small, 0.5 moderate, 0.8 large; Cohen, 1988]), and the improvement index was +11. Improvement Index is the expected change in percentile rank for an average non-PAT comparison group student if the student had received the PAT intervention. It is measured as the percentile difference between the PAT intervention group mean and the non-PAT comparison group mean using the non-PAT comparison group distribution. The resulting improvement index is +11, corresponding to moving performance for the average student from the 50th to the 61st percentile of the comparison group distribution (WWC, 2017).

Exhibit 8. AzMERIT Math Mean Scores for PAT and Non-PAT Comparison Group Students, 2015 to 2017

Group	N	2015 AzMERIT Math Scale Score Mean (SD)	2017 AzMERIT Math Scale Score Mean (SD)	Gain Score	t	p	ES	Improvement Index
PAT	364	3567.07 (50.15)	3623.97 (46.89)	56.90	4.667	**.000**	.28	+11
Non-PAT	1193	3584.75 (57.28)	3632.93 (50.82)	48.18				

Note: Gain Score = difference from 2015 to 2017 mean math scale score. ES = Effect size. Improvement Index = the expected change in percentile rank for an average comparison group student if the student had received the intervention. It is measured as the percentile difference between the intervention group mean and the comparison group mean using the comparison group distribution. P-values are deemed significant if ≤ .05. Significant p-values are shown in bold font.

Reading Achievement – RAPS 360

Is there a difference in reading outcomes in fluency between PAT and non-PAT students as measured by RAPS 360? PAT participants showed larger mean gains from 2015 to 2017 and were associated with a larger adjusted 2017 mean score than the non-PAT comparison students after controlling for 2015 mean scores. However, the observed difference was not statistically significant (see Exhibit 8). Baseline equivalence on RAPS 360 Fluency scale scores showed that the difference in 2015 means of the two groups was 6% (.06) of the PSD. ANCOVA was conducted between the PAT group (n = 362) and the non-PAT comparison group (n = 932) and did not show a statistically significant difference in the adjusted posttest means between the PAT participant's (M = 131.00; SD = 58.52) and non-PAT comparisons (M = 126.55; SD = 58.80), after controlling for the 2015 covariate. The fixed factor 'group' was not significant, $F(1, 1291)$ = 2.03, p = .155, however the covariate 2015 was significant, $F(1, 1293)$ = 456.45, p < .001. The homogeneity of variance assumption held with a Levene's test of .886. Hedges' g effect size calculation = .08. Although PAT participants were associated with a larger adjusted 2017 mean score than the non-PAT comparison students after controlling for 2015 scores, the difference was not statistically significant.

Exhibit 9. RAPS 360 Fluency Scale Scores for PAT and Non-PAT Comparison Group Students, 2015 to 2017

Group	2015 RAPS 360 Fluency Scale Scores	2017 RAPS 360 Fluency Scale Scores				
	N	*Mean (SD)*	*Mean (SD)*	*F*	*p*	*ES*	Improvement Index
PAT	362	83.88 (42.15)	131.00 (58.52)	2.03	.15	.08	+3
Non-PAT	932	86.41 (42.00)	126.55 (58.80)				

Note. ES = the adjusted posttest means and unadjusted standard deviations were used to compute Hedges' g effect size (.08), p = .155. Improvement Index = the expected change in percentile rank for an average comparison group student if the student had received the intervention. P-values are deemed significant if ≤ .05.

Non-parametric Mann-Whitney U tests were also conducted on grade-level scores on the RAPS 360 Phonics, Comprehension, Phonemic Awareness and Listening/Vocabulary subscales to test for differences in mean ranks between the PAT and non-PAT comparison groups at the end of the spring school year in 2015, 2016, and 2017. Exhibit 10 shows the results of this analysis and statistically significant p-values shown in bold. Significant findings include:

- **Spring 2015**: Scores in Phonics were significantly higher for PAT participants (Mean Rank = 1605.93) than for non-PAT comparisons (Mean Rank = 1496.16), U = 716784.00, p = .004. Effect size r ($z / \sqrt{N}$) = .05.
- **Spring 2016**: Scores in Phonics were significantly higher for PAT participants (Mean Rank = 1603.25) than for non-PAT comparisons (Mean Rank = 1460.09), U = 671456.00, p < .001. Effect size r ($z / \sqrt{N}$) = .07.

- **Spring 2017:** Scores in Phonics were significantly higher for PAT participants (Mean Rank = 1616.54) than for non-PAT comparisons (Mean Rank = 1432.55), $U = 638233.00$, $p < .001$. Effect size r ($Z / \sqrt{N}$) = .09.
- **Spring 2017:** Scores in Comprehension were significantly higher for PAT participants (Mean Rank = 1195.77) than for non-PAT comparisons (Mean Rank = 1112.47), $U = 638233.00$, $p <$ 001. Effect size r ($Z / \sqrt{N}$) = .05.

The other subscales of Phonemic Awareness and Listening/Vocabulary Levels showed no significant difference between the PAT and non-PAT comparison group.

Exhibit 10. RAPS 360 Grade Level Scores on RAPS 360 Subscales for PAT and Non-PAT Comparison Group Students, 2015 to 2017

Year	RAPS 360 Subscales	Group	N	Mean Rank	U	p
Spring 2015	Phonics	PAT	646	1605.93	716784.00	**.004**
		Non-PAT	2392	1496.16		
	Comprehension	PAT	553	1198.51	509595.00	.141
		Non-PAT	1921	1248.72		
	Phonemic Awareness	PAT	107	343.36	27460.00	.310
		Non-PAT	546	323.79		
	Listening/ Vocabulary Level	PAT	597	1333.22	617426.50	.647
		Non-PAT	2094	1349.64		
Spring 2016	Phonics	PAT	633	1603.25	671456.00	**.000**
		Non-PAT	2347	1460.09		
	Comprehension	PAT	537	1268.00	483836.00	.064
		Non-PAT	1900	1205.15		
	Phonemic Awareness	PAT	121	337.78	30397.50	.393
		Non-PAT	528	322.07		
	Listening/ Vocabulary Level	PAT	580	1328.98	561740.00	.198
		Non-PAT	2007	1283.89		
Spring 2017	Phonics	PAT	631	1616.54	638233.00	**.000**
		Non-PAT	2312	1432.55		
	Comprehension	PAT	503	1195.77	409557.50	**.011**
		Non-PAT	1758	1112.47		
	Phonemic Awareness	PAT	138	320.70	30332.00	.215
		Non-PAT	471	300.40		
	Listening/ Vocabulary Level	PAT	567	1347.36	554887.50	.128
		Non-PAT	2042	1293.24		

Note. U = the Mann Whitney U value that tests for differences in mean ranks between the PAT and non-PAT comparison groups. P-values are deemed significant if ≤ .05. Significant p-values are shown in bold font.

English Language Learner Achievement - AZELLA

Is there a difference in reading outcomes between PAT and non-PAT students as measured by the AZELLA? The results suggest PAT ELL students performed better than non-PAT ELL students on the AZELLA. PAT participation was associated with consistently larger mean scores compared to non-PAT participation in reading, writing, and total scores for all three years and a statistically significant MANOVA effect was obtained, as six of the nine *F*-tests were statistically significant.

A one-way Multivariate Analysis of Variance (MANOVA) was conducted to test the hypothesis that there would be one or more mean differences between PAT students (n = 60) and non-PAT comparison students (n = 229) on AZELLA assessment outcomes in reading, writing, and total combined scores over three years (2015, 2016, and 2017). Prior to conducting the MANOVA, a series of Pearson correlations were performed between the dependent variables to test the assumption that the dependent variables would be correlated in the moderate range (i.e., 0.20 - 0.60; Meyers et al., 2006) (see Exhibit 11). A meaningful pattern of correlations was observed among most of the dependent variables, suggesting the appropriateness of a MANOVA. Additionally, the Box's M value of 113.95 was associated with a p value of 0.006, which was interpreted as non-significant based on Huberty and Petoskey's (2000) guideline (i.e., p <0.005). Thus, the covariance matrices between the groups were assumed to be equal for the purposes of the MANOVA.

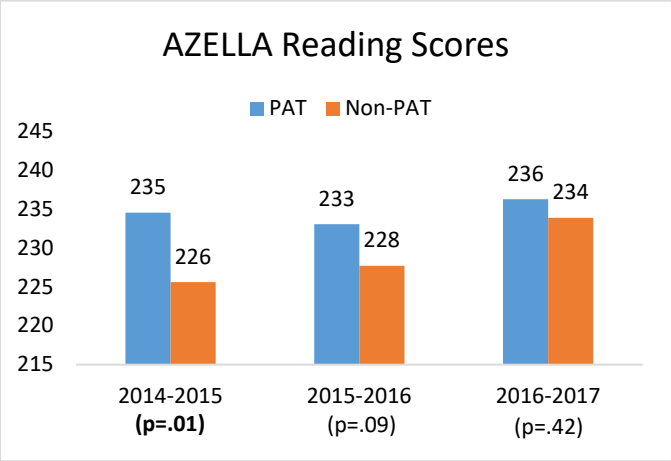
Exhibit 11. Pearson Correlations, Means, and Standard Deviations Associated with AZELLA Reading, Writing and Total Scores

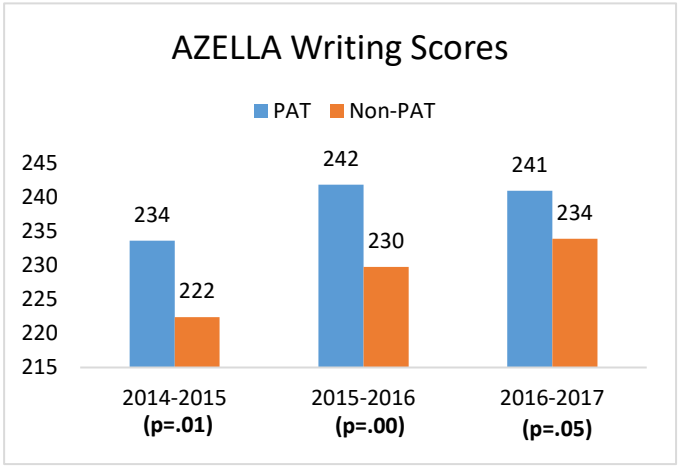
	1.	2.	3.	4.	5.	6.	7.	8.	9.	*M*	*SD*
1. Read '15	1.0									227.45	24.34
2. Write '15	.64	1.0								224.75	29.72
3. Total '15	.59	.65	1.0							2379.91	60.69
4. Read '16	.42	.38	.25	1.0						228.82	22.02
5. Write '16	.37	.49	.40	.67	1.0					232.30	25.39
6. Total '16	.24	.46	.71	.54	.67	1.0				2413.28	48.05
7. Read '17	.30	.28	.21	.39	.35	.18	1.0			234.36	20.53
8. Write '17	.34	.30	.17	.32	.37	.28	.66	1.0		235.38	24.40
9. Total '17	.23	.40	.68	.26	.46	.80	.48	.53	1.0	2444.69	50.18

Note. N = 692; Correlations greater than .10 are statistically significant (p < .01).

As shown in Exhibits 12 and 13, PAT students had significantly higher mean scores than the non-PAT comparison group students for six of the nine *F*-tests: Reading in 2015, Writing in 2015, 2016, and 2017, and Total Scores in 2015 and 2016. Effect sizes for the *F*-tests were small ranging from 0.002 to a high of 0.037. A statistically significant MANOVA was observed (Wilks' Lambda = 0.94; $F(9, 279) = 2.08$; $p = 0.03$) and the multivariate effect size was estimated at 0.063. This finding implies that 6.3% of the variance in the canonically derived dependent variable was accounted for by the group variable (PAT/non-PAT comparisons).

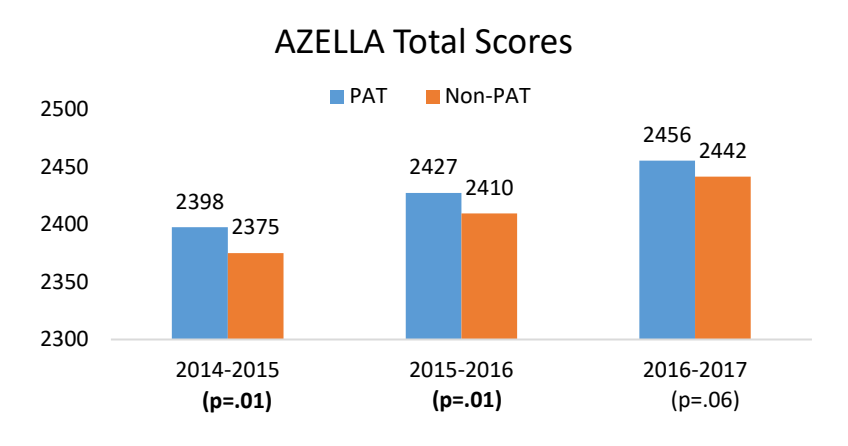
Exhibit 12. AZELLA Reading and Writing Scores for PAT and Non-PAT Comparison Group Students





Note. PAT participants (n=60); Non-PAT comparisons (n=229). P-values are deemed significant if ≤ .05. Significant p-values are shown in bold font.

Exhibit 13. AZELLA Total Scores for PAT and Non-PAT Comparison Group Students



Note. PAT participants (n=60); Non-PAT comparisons (n=229). P-values are deemed significant if ≤ .05. Significant p-values are shown in bold font.

Absenteeism and Suspension Outcomes

Is there a difference between PAT and non-PAT students on child school indicators of absence rates and number of suspensions (in-school and out-of-school) in the last few school years?

Absenteeism

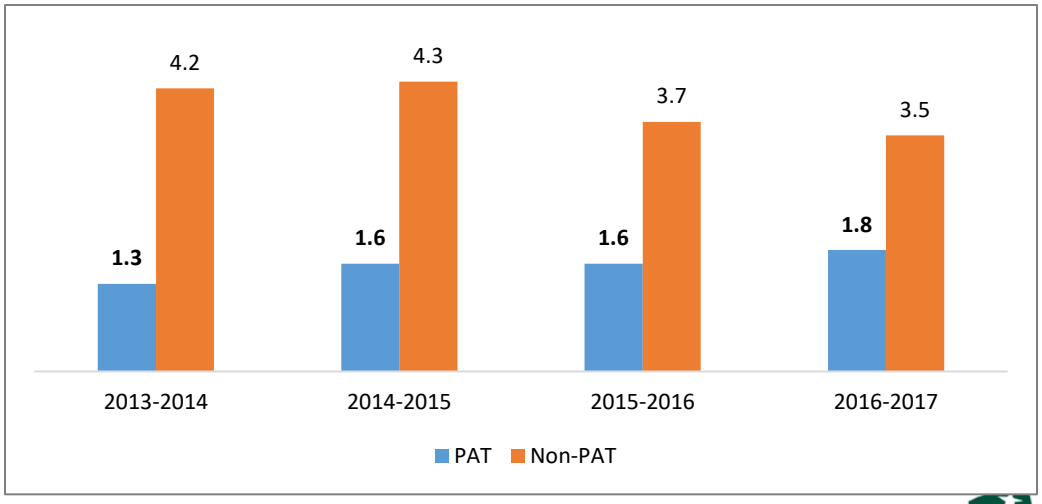
Comparing absence rates of PAT students and non-PAT comparison students, Exhibits 14 and 15 show that PAT students had significantly lower average rates of absenteeism than non-PAT comparison students over four school years, from 2014 to 2017 (p = .000). Independent samples t-tests showed that PAT students had significantly lower average absence rates, ranging from an average of 1.3 days to 1.8 days, compared to non-PAT students, with significantly higher average absence rate ranging from 3.5 days to 4.3 days. The effect size for each school year ranges from a low of .28 to a high of .49, suggesting that participation in the PAT program produces a small to medium effect on student absenteeism.

Exhibit 14. Average Absence Rates by PAT and Non-PAT Comparison Group Students, 2014 to 2017

School Year	Group	*N*	Mean Absence Rate	SD	*t*	*p*	*ES*
2013-2014	PAT	671	1.3	3.4	14.638	**.000**	.49
	Non-PAT	2822	4.2	7.7			
2014-2015	PAT	670	1.6	3.7	13.183	**.000**	.43
	Non-PAT	3057	4.3	8.1			
2015-2016	PAT	705	1.6	3.6	10.564	**.000**	.32
	Non-PAT	3305	3.7	8.5			
2016-2017	PAT	745	1.8	3.8	8.554	**.000**	.28
	Non-PAT	3307	3.5	7.8			

Note. ES = effect size for independent samples T-test is Cohen's *d*. P-values are deemed significant if ≤ .05. Significant p-values are shown in bold font.

Exhibit 15. Bar Chart of Average Absence Rates by PAT and Non-PAT Comparison Group Students, 2014 to 2017



Suspensions

In-School Suspensions

Exhibit 16 shows the results of a chi square test comparing PAT students with non-PAT matched comparison students across categories of in school suspensions for three school years. In general, the majority of students in both groups had zero in-school suspension days (ranging from 96% to 99% for both groups); a small percentage had 1 to 3 in-school suspension days (ranging from 1% to 3% for both groups); and an even smaller percentage had 4 or more in-school suspension days (ranging from .2% to .7% for both groups). Significance was observed favoring the PAT group over the non-PAT matched comparison group only for 2014-2015 (p = .03). Significance was not observed for the other school years assessed.

Exhibit 16. In-School Suspension Days by PAT and Non-PAT Comparison Students, 2015 to 2017

School Year	Group	In-School Suspension Day Categories					
		0 Days % (*n*)	*1–3 Days* % (*n*)	*4+ Days* % (*n*)	x^2	*p*	*ES*
2014-2015	PAT	99% (973)	1% (6)	.4% (4)	6.845	**.03**	.03
	Non-PAT	98% (4805)	2% (84)	.5% (26)			
2015-2016	PAT	98% (967)	1% (14)	.2% (2)	4.894	.09	.03
	Non-PAT	97% (4780)	2% (102)	.7% (33)			
2016-2017	PAT	97% (957)	2% (22)	.4% (4)	2.428	.29	.02
	Non-PAT	96% (4736)	3% (151)	.6% (28)			

Note: PAT participants (n = 983); Non-PAT comparisons (n = 4915). The observed count is shown, however, none of the cells produced an expected count less than 5. The minimum expected count to run a valid Chi Square test is n = 5. ES = effect size for Chi Square is Cramer's V. Within groups, percentages may total over 100% due to rounding. P-values are deemed significant if ≤ .05. Significant p-values are shown in bold font.

Out-of-School Suspensions

Similar to in-school suspensions, Exhibit 17 shows that the majority of both PAT and non-PAT matched comparison students had zero out-of-school suspension days. A Chi Square test produced significant results favoring the PAT group over the non-PAT group for 2014-2015 (p = .03) only. Significance was not observed for the other school years assessed.

Exhibit 17. Categories of Out-of-School Suspension Days by PAT and Non-PAT Comparison Group Students, 2015 to 2017

School Year	Group	Out-of-School Suspension Day Categories					
		0 Days % (*n*)	*1–3 Days* % (*n*)	*4+ Days* % (*n*)	x^2	*p*	*ES*
2014-2015	PAT	98% (967)	1% (14)	.2% (2)	7.030	**.03**	.04
	Non-PAT	98% (4796)	1% (65)	1% (54)			
2015-2016	PAT	98% (961)	1% (12)	1% (10)	4.210	.24	.02
	Non-PAT	97% (4745)	2% (86)	2% (84			
2016-2017	PAT	97% (957)	2% (18)	1% (8)	3.449	.18	.02
	Non-PAT	96% (4736)	2% (103)	2% (76)			

Note. PAT participants (n = 983); Non-PAT comparisons (n = 4915). The observed count is shown, however, none of the cells produced an expected count less than 5. The minimum expected count to run a valid Chi Square test is n = 5. ES = effect size for Chi Square is Cramer's V. Within groups, percentages may total over 100% due to rounding. P-values are deemed significant if ≤ .05. Significant p-values are shown in bold font.

Parenting Outcomes

To what extent does the PAT intervention improve parenting skills as evaluated by testing the hypothesis that pretest and posttest means are equal in a within group analysis of parents as measured by KIPS and the PFS?

Keys to Interactive Parenting Scales

To test the null hypothesis that pretest and posttest mean KIPS scores were equal among the same group of PAT parents, a dependent samples *t*-test was performed (see Exhibit 16). The null hypothesis of equal means was rejected ($t(181) = 5.008$; $p < .001$). PAT parents demonstrated a significant improvement in parenting quality over time from pretest and posttest KIPS assessment. Cohen's *d* was estimated at .37, which is a medium effect based on Cohen's (1988) guidelines.

Exhibit 18. Change in PAT Parents' Total Average KIPS Pretest and Posttest Scores

Assessment	*Mean Score (SD)*	Diff	*t*	*p*	ES	N
Pretest Total Mean Score	3.66 (.61)	.21	5.008	**.000**	.37	182
Posttest Total Mean Score	3.87 (.71)					

Note. Diff = difference from 2015 to 2017 math scale score. ES = Cohen's *d* is the calculated effect size for a paired-sample t-test. P-values are deemed significant if ≤ .05. Significant p-values are shown in bold font.

Protective Factors Survey

A dependent samples *t*-test was also performed on the four PFS subscale scores to test for equal mean scores for PAT parents from pretest to posttest (see Exhibit 17). The null hypothesis of equal means was rejected for three of the four subscales. PAT parents showed significant improvements over time in the subscale domains of family functioning, social support, and concrete support from pretest to posttest.

Exhibit 19. Change in PAT Parents' PFS Subscale Pretest and Posttest Scores

PFS Subscale	Pretest Mean Score (SD)	Posttest Mean Score (SD)	Diff	*t*	*p*	ES	N
Family Functioning/Resiliency Subscale	6.10 (.89)	6.29 (.87)	.18	2.853	**.005**	.18	250
Social Support Subscale	6.25 (.98)	6.42 (.83)	.18	2.796	**.006**	.18	250
Concrete Support Subscale	4.01 (2.05)	4.91 (2.06)	.91	5.311	**.000**	.34	248
Nurturing and Attachment Subscale	6.69 (.38)	6.74 (.37)	.05	1.699	.091	.11	250

Note. Diff = difference from 2015 to 2017 math scale score. ES = Cohen's *d* is the calculated effect size for a paired-sample t-test. P-values are deemed significant if ≤ .05. Significant p-values are shown in bold font.

Discussion

The aim of this study was to estimate the effect of the PAT home visiting intervention on SUSD students and their parents who participated in the program (i.e., PAT group), relative to a comparison group of SUSD students who did not participate in the PAT program (non-PAT group). Using a matched comparison group quasi-experimental design, this study assessed three years of academic and school outcomes in the areas of:

- Reading and math standardized assessment scores;
- Achievement of English language proficiency of ELL students; and
- School indicators of absentee rates and number of suspensions.

This study also examined change in parenting skills of parents who participated in the PAT program from baseline to posttest assessment (within group change) on two parenting measures: KIPS and PFS. Overall, findings from this study indicate that the PAT home visiting program was associated with statistically significant improvements in child academic outcomes (e.g., ELA/reading achievement, math, ELL English proficiency), and parenting skills. The current study extended past research on PAT by also ascertaining that PAT participation was associated with statistically significant decreases in absenteeism and in- and out-of-school-suspensions.

PAT participants displayed statistically significant higher ELA/reading achievement scores relative to the non-PAT comparison group as measured by the AzMERIT (Hedges' g = 0.10); although this effect size is considered small (Cohen, 1988), it was the equivalent of moving a child from the 50th percentile to the 54th percentile in ELA/reading achievement. In terms of math, PAT participation was associated with statistically significant higher math gain scores for the PAT group as compared to the non-PAT group, as measured the by AzMERIT (Hedges' g = 0.28). These findings suggest that the PAT program played a role in the improvements in ELA/reading and math achievement displayed by the PAT student group. While the relationship between parenting skills and academic outcomes was not tested, because the PAT program focuses mainly on increasing parental knowledge of child development and improving parenting skills, significant improvements observed in parenting skills likely played a role in the improved academic achievement outcomes displayed by PAT children.

It is interesting and somewhat enigmatic that although PAT participants demonstrated significantly higher AzMERIT ELA/reading achievement scores relative to the non-PAT group, the increase in RAP-360 scores did not reach statistical significance. While the RAP-360 scores did improve in PAT participants relative to non-PAT participants, this improvement was only statistically significant for the phonics subscale and the comprehension subscale in SY2017. Perhaps PAT participation is associated with improved general reading achievement as measured by the AzMERIT, but PAT participation might not affect the specific reading skills measured by the RAP-360.

It is also noteworthy that the PAT group had statistically significant higher scores in English language ability for ELL students as measured by the AZELLA as compared to the non-PAT group. Providing PAT for ELL students appears to be particularly important given their increased risk for negative academic outcomes. Providing PAT to this vulnerable group was associated with higher scores on a test of English language ability, which could serve to keep ELL students engaged in school, a vital protective factor for future positive developmental outcomes.

PAT participation was also significantly associated with lower absenteeism across all school years examined, as well as in- and out-of-school-suspensions in SY 2015. This lower rate of absences could indicate that PAT parents were more aware of the importance of education because of PAT participation and therefore were more inclined to ensure that their children attended school. The lower rates of school suspensions suggest that PAT participation was associated with improved child behavior. This finding highlights the notion that altering and improving parenting behavior and skills could in fact result in improved child behavior as well. However, mediational analysis between parenting quality and student outcome, thus further research on the causal links between improved PAT parenting behavior and child behavior is needed.

Finally, findings of the current study show that PAT participation was associated with statistically significantly higher parenting quality, as measured by the KIPS and PFS domains of family functioning, social supports, and concrete supports. The effect sizes were moderate (KIPS [ES = 0.48] PFS [ES = 0.18, .18, .34]; Cohen, 1988), but quite impressive for intervention research. This finding highlights the success of the PAT curriculum in improving parenting skills, knowledge, and confidence. Based on these results, it appears that the PAT curriculum provides parents with the necessary skills and knowledge to improve their parenting. Future research should investigate and confirm that this improved parenting is the mechanism of change that is responsible for improvements in child academic and behavioral outcomes.

Conclusion

Findings of the current study suggest that participation in the PAT program was associated with statistically significantly higher scores in student level outcomes including ELA/reading achievement, math achievement, and ELL student English language proficiency, as well as statistically significantly lower absenteeism and school suspensions. For parents, PAT participation was associated with statistically significantly higher parenting knowledge of child development, parent nurturance, parent-child attachment, family functioning, family social support, parent confidence, and parenting skills. These findings highlight the utility of the PAT home-visitation program in improving the family Microsystem, which likely helped improve children's academic achievement outcomes. Overall, current findings reinforce the importance

of early and ongoing supportive interventions in terms of improving family functioning and child developmental outcomes.

Study Limitations

Findings of the current study must be understood in light of specific limitations. First, the lack of random assignment among treatments and comparisons leads to non-equivalent test groups which can reduce internal validity, and conclusions about causality are less definitive in quasi-experimental designs. Straight comparisons of outcomes between PAT participants and a comparison group do not meet the requirements for making a clear causal inference. For these reasons, it was imperative that the two groups were well-matched on key variables of gender, ethnicity, special education status, ELL and free/reduced lunch status. This is important because the more similar the two groups are at baseline, the more observed differences between the two groups could be attributed to the intervention. Yet, even though Propensity Score matching was successfully applied in identifying similar groups on key demographics mentioned above, the inability to rule out the possibility that outcomes are unrelated to program participation still exist.

Second, the initial data on children as PAT participants for this study were contained in two files: a file of 1,443 students that were former PAT participants, with birthdates ranging from September 1999 through October of 2010, and a file of 642 children pre-Kindergarten through 3rd grade with birthdates ranging from September 2008 to June 2017, who are receiving or had participated in PAT services during early childhood. The overlap in birthdates from these two files with participants born from 9/2008 through 10/2010, captured a very small number of PAT participants in the group variable. This effected the ability to provide PAT dosage analysis, as data on the frequency of home visits are available for children who participated in the PAT intervention after 2011, but not before. A within group analysis - within the PAT group only, would have allowed us to address dosage-related research questions such as the relationship between the duration of PAT participation with their school achievement outcomes, or the relationship between frequency of home visits and various school outcomes. Future research should examine the impact of program dosage on child and parent outcomes.

Third, it would have been ideal to have a control group of non-PAT parents in order to compare changes in KIPS and PFS scores to the PAT parent group. Additionally, the KIPS measure was completed by PAT staff without assessment of interrater reliability and may include bias as staff make observations of families on a rating scale and are likely to believe their efforts are leading to positive outcomes.

Fourth, our entire sample came from Arizona and the majority of the sample was Latino/Hispanic. These unique sample characteristics indicate that findings should be generalized to other geographic regions and races with caution. Future PAT research should use more geographically and ethnically/racially diverse samples in order to better understand how the program might operate differently across populations.

Finally, outcome data on children's academic performance in elementary school is limited until more PAT children enter the school system and take part in assessments. This lack of connection (from limited passage of time) between dosage data (post-2011) and PAT children assessed in the public-school system produced very little overlap between PAT participation as late as 2011 and, having been in the school district long enough to establish records on their academic measures. Yet, the improvements in data collection and maintenance starting in 2011 by PAT staff create a promising opportunity for future analyses, as PAT students move through the school district and complete their academic assessments.

References

Arizona Department of Education (ADE), & American Institutes for Research. (2017). *Annual Technical Report: Arizona Statewide Assessment in English Language Arts and Math 2015-2016 School Year*. Phoenix, AZ: Author.

Arizona Department of Education & Harcourt Assessment, Inc. (2007). *Arizona English Language Learner Assessment: Technical manual*. Phoenix, AZ: Author.

Arizona Department of Education. (2014). *Arizona English Language Learner Assessment (AZELLA): Evidence-Based Standard Setting Review Report Kindergarten Placement Test.* Phoenix, AZ: Author.

Arizona Department of Education (ADE). (2016). *Arizona English Language Learner Assessment: 2016 Technical Report.* Phoenix, AZ: Author.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 12*, 155-159.

Comfort, M., & Gordon, P. R. (2006). The Keys to Interactive Parenting Scale (KIPS): A practical observational assessment of parenting behavior. *NHSA Dialog: A Research-To-Practice Journal for the Early Intervention Field, 9*(1), 22–48.

Comfort, M., Gordon, P. R., English, B., Hacker, K., Hembree, R., Knight, C. R., & Miller, C. (2010). Keys to Interactive Parenting Scale: KIPS shows how parents grow. *Zero to Three Journal, 30*(4), 33–39.

Comfort, M., Gordon, P.R., & Naples, D. (2011). KIPS: An evidence-based tool for assessing parenting strengths and needs in diverse families. *Infants & Young Children, 24*(1), 56-74.

Comfort, M., Gordon, P. R., & Unger, D. G. (2006). Keys to Interactive Parenting Scale: A window into many facets of parenting. *Zero to Three Journal, 26*(5), 37–44.

Counts, J.M., Buffington, E.S., Chang-Rios, K., Rasmussen, H.N., Preacher, K.J. (2010). The development and validation of the protective factors survey: A self-report measure of protective factors against child maltreatment. *Child Abuse and Neglect*, *34*, (10) 762-772.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*. 16, 297-334.

Ellis, P. D. (2010). *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*. New York: Cambridge University Press. pp. 9-11.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007) Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, *15*(3), 199-236.

Huberty, C. J., & Petoskey, M. D. (2000). Multivariate analysis of variance and covariance. In H. Tinsley and S. Brown (Eds.) *Handbook of applied multivariate statistics and mathematical modeling*. New York: Academic Press.

Meyers, L. S., Gamst, G., & Guarino, A. (2006). *Applied multivariate research: Design and interpretation*. Thousand Oaks, CA: Sage Publishers.

Mindplay & Methods and Solutions, Inc. (2015). *Universal Screener RAPS 360: Reading Analysis & Prescription System*. Tucson, AZ: Author.

Nunnaly, J. (1978). *Psychometric Theory*. New York: McGraw-Hill.

Parents as Teachers. (2018). *Quality Assurance Guidelines for Parents as Teachers Affiliates*. St. Louis, MO: Author.

Parents as Teachers National Center. (2014). *Foundational 2 Curriculum: 3 years through kindergarten*. St. Louis, MO: Author.

Parents as Teachers National Center. (2015). *Parents as Teachers Foundational Training Guide*. St. Louis, MO: Author.

Parents as Teachers National Center. (2017). *Foundational Curriculum*. St. Louis, MO: Author.

Posten, H. O. (1984). Robustness of the two-sample t-test. In D. Rasch and M.L. Tiku (Eds.), *Robustness of statistical methods and nonparametric statistics* (pp. 92 – 99). Dordrecht, Germany: Reidel.

Randolph, J. J., Falbe, K., Manuel, A. K., & Balloun, J. L. (2014). A step-by-step guide to propensity score matching in R. *Practical Assessment, Research & Evaluation, 19*(18).

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse (2017). WWC Procedures and Standards Handbook, v4. Available at: https://ies.ed.gov/ncee/wwc/Handbooks.