Senior Project

Department of Economics

The Effect of Subsidized Pre-Kindergarten on Poverty Rates

Nathan D. Riley

Spring 2024

Advisor: *Dr. Ali Ena*

I. Abstract

This paper explores the effects of subsidized Pre-K programs on poverty rates across the United States using data from the National Institute for Early Education Research (NIEER) and the U.S. Census Bureau. The study leverages a two-way fixed effects model (TWFE) to analyze how funding, enrollment, and program quality influence state-level poverty rates over a 21-year period. Key findings suggest that increased state spending per child in Pre-K programs correlates with reduced poverty rates, while total spending per child has a mixed impact. Increased enrollment rates for 3-year-olds show a statistically insignificant association with poverty reduction, and similar patterns are found for 4-year-old enrollment rates. Additionally, quality standards met through Pre-K programs contribute minimally to poverty reduction. This comprehensive study contributes to the literature by analyzing multiple variables influencing the success of state-funded Pre-K programs at a national level, providing insights into how spending, enrollment, and program quality interact. The findings suggest that state funding impacts the efficacy of these programs modestly, whereas total spending and quality benchmarks require further investigation to clarify their roles. Future research is encouraged to delve deeper into individual-level data to refine controls for parental workforce participation and demographics like age and gender. Incorporating localized geographic fixed effects could uncover regional disparities in program implementation and identify inequities. By providing evidence of the correlations between Pre-K funding and poverty alleviation, the study offers policymakers a data-driven framework to optimize early education policies for maximum socioeconomic impact.

Table of Contents

II. Introduction

The study of state-funded Pre-K programs' impact on poverty rates is crucial for economists, policy makers, and the global public, as it provides essential data for optimizing educational policies and offers a deeper understanding of how early education can be a pivotal tool for economic and social advancement worldwide. This research not only helps in crafting effective strategies to alleviate poverty but also enriches our understanding of the broader economic and sociological effects of Pre-K education.

The exploration of state-funded Pre-K programs' effects on poverty rates offers valuable insights for policy makers, who can utilize this data to optimize the structure and spending of these educational initiatives. Understanding the key components that contribute to the success of Pre-K programs aids in crafting policies that effectively combat poverty and stimulate economic development.

This research also holds global significance outside of the United States, as education serves as a crucial means for individuals worldwide to rise out of poverty. By analyzing the impact of Pre-K programs, insights can be drawn that highlight both the strengths and weaknesses of current educational models. This enables citizens around the world to better advocate for educational policies that address significant factors influencing their effectiveness.

Economists are provided with an opportunity to delve into the complex interactions between Pre-K education and poverty alleviation. Such studies can reveal how early childhood education influences other important economic and sociological variables, broadening the scope of research into the socioeconomic impacts of Pre-K programs and informing future economic analyses.

In contrast to previous analyses that were confined to smaller scales with more granularity (Rossin-Slater & Wüst, 2020; Pearman, 2020; Martinez et al., 2017), this study takes a broader approach to explore the effects of Pre-K programs on poverty rates across the entire United States. By extending the scope to a national level and considering attributes relevant to Pre-K programs, the research makes a contribution to the existing body of knowledge in early childhood education. Moreso, the connection between early childhood education and poverty alleviation is largely unexplored in existing research. This novel approach facilitates a more thorough understanding of the relationship between Pre-K programs and poverty outcomes, offering insights for policymakers seeking effective strategies.

One differentiating aspect of this research lies in its analysis of Pre-K attributes such as spending, enrollment, and quality. This examination aims to discern which metrics bear significant relevance to poverty reduction and which may not. By dissecting program quality into various measurable components, the study provides an understanding of how different aspects contribute to the overall impact on poverty rates. These details enable policymakers to tailor interventions and improvements to specific areas that prove most influential in mitigating the effects of poverty through Pre-K programs, fostering a more targeted and effective approach to early childhood education policy.

In the following sections, I unfold the layers surrounding the impact of state-funded Pre-K programs on poverty rates. The Literature Review section critically examines the existing dialogue, highlighting the necessity for an expansive national-level analysis. Subsequently, the Theory and Methodology section outlines the analytical frameworks and methods employed, emphasizing the TWFE analysis approach. The Results section presents the findings from the investigation and a framework for interpretation.

III. Literature Review

There is a notable gap in the existing literature concerning comprehensive examination of how Pre-K quality, spending, and enrollment collectively contribute to poverty rates in the United States. Despite valuable insights offered by previous research on different facets of state-funded Pre-K programs, existing studies, such as those conducted by Rossin-Slater & Wüst (2020), Pearman (2020), and Martinez et al. (2017), are limited to smaller scales, emphasizing the necessity for comprehensive research on a larger, more expansive level, at the expense of granularity. Previous studies have often focused on isolated aspects, lacking a unified understanding of how spending, quality, and enrollment can have an impact on the efficacy of Pre-K programs in their respective outcome interests, in the case of this study, poverty.

Longitudinal studies, such as those conducted with the Abecedarian project (Campbell et al., 2002), Child-Parent Center (CPC) (Reynolds, 1994), and Perry Preschool (Manning & Patterson, 2006; Schweinhart et al., 2005), have revealed substantial and lasting effects of early education programs on inequality. These effects manifest through various factors, including heightened social intelligence, success in school, and increased educational attainment. Among these studies, the CPC project, which involved 1,106 low-income African American children, stands out as the largest. This particular study additionally finds a significant association between program participation and positive outcomes such as increased parental involvement and a decrease in special education placements, aligning with expectations.

Crucially, there is a distinct lack of any large-scale studies exploring the effects of subsidized Pre-K on poverty rates at all, even removing the qualifier that they also consider variables other than when a program started. This gap becomes more apparent when contrasted with similar research, such as Bartik & Hershbein (2018), which focuses solely on the relationship between

Pre-K programs and test scores. Notably, their study deemed only five programs at the time to be of sufficient quality to be considered 'universal' and therefore included in the treatment group. In contrast, my research aims to adopt a more inclusive approach, incorporating all state-funded Pre-K programs, but taking into consideration the aforementioned variables. Additionally, this study intends to explore modifiers that may influence the success of these programs, contributing to a more comprehensive understanding of their impact on poverty rates. This broader perspective is crucial for informing evidence-based policies and addressing the multifaceted challenges posed by poverty on a national scale.

Within the context of this research, three critical dimensions emerge as pivotal aspects in understanding the relationship between state-funded Pre-K programs and poverty rates: Spending, Quality, Enrollment Rates, and demographic attributes. Due to the limitations of current publicly accessible data, it is not realistic control for demographic attributes.

Spending, specifically the financial resources allocated per pupil to subsidized Pre-K programs, stands as the cornerstone influencing both program quality and accessibility. Adequate financial support is not only essential for implementing high-quality educational initiatives but also directly impacts the breadth and depth of program accessibility. Recognizing that quality programs cannot be sustained without adequate spending, analyzing spending per pupil becomes paramount in understanding how financial commitments set the stage for the overall success and impact of state-funded Pre-K initiatives (Freiman-Krauss et. al., 2023; Lamy, 2013).

The significance of spending extends beyond the initial stages of program implementation. Continued financial support is vital for maintaining program quality over time and adapting to evolving educational needs. Inadequate spending can lead to challenges such as limited teacher training, outdated learning resources, and inadequate facilities, ultimately hindering the program's effectiveness. Additionally, equitable distribution of funds is crucial to ensure that all children, irrespective of socio-economic backgrounds, have access to a high-quality Pre-K education (McCoy et al., 2015).

The quality of state-funded Pre-K programs is pivotal, directly impacting educational outcomes and, consequently, their potential influence on poverty rates. High-quality programs are characterized by effective curriculum design, qualified teaching staff, and robust program infrastructure. Analyzing program quality enables a nuanced understanding of how these attributes contribute to positive cognitive and socio-emotional outcomes for participants, forming a critical link in the chain between spending, program quality, and their collective impact on poverty rates (Friedman-Krauss et. al., 2023; Barnett, 2003).

Enrollment rates in state-funded Pre-K programs serve as key indicators of accessibility and participation levels, crucial for understanding the programs' overall reach and effectiveness. Examining enrollment patterns provides insights into disparities in access among diverse demographic groups. Accessibility, as indicated by enrollment rates, is a crucial aspect in evaluating how state-funded Pre-K programs, underpinned by adequate spending and quality, contribute to mitigating poverty rates on a broader, national scale (Friedman-Krauss et. al., 2023).

Understanding the impact of subsidized Pre-K programs on poverty rates necessitates grappling with neighborhood disparities, a challenge amplified in a national context. While studies within individual states adeptly navigate intra-state variations, broader analyses are limited to capturing inter-state differences. This limitation becomes pronounced as programs like the Tennessee Voluntary Prekindergarten (TN-VPK) Program, designed to target impoverished children, exhibit variations in quality influenced by neighborhood income levels (Valentino, 2018; McCoy et al., 2015).

Despite TN-VPK's intended focus on impoverished children, disparities exist in program quality, negatively affecting children in disadvantaged areas. The negative correlation between neighborhood poverty and Pre-K program quality underscores the need to address these complexities. This is crucial, as studies suggest that Pre-K programs affect low-income children more significantly than their middle and high-income counterparts. While our study captures inter-state variations, recognizing the existence of these disparities at the national level is important for a comprehensive understanding of how state-funded Pre-K programs, even those targeting impoverished children, contribute to shaping poverty rates. Unfortunately with current data a nation-wide analysis cannot address intra-state treatment disparities experienced, most notably between high- and low-income areas.

IV. Data

The data for this analysis is sourced primarily from two reputable organizations: NIEER, the premier early education research organization based out of Rutgers and the United States Census Bureau. The dataset encompasses an aggregation of state-year information, covering all 50 states in the United States and the District of Columbia. The temporal scope of the data spans from the academic years 2002-03 to 2022-23, totaling 21 years and resulting in a dataset with a sample size (N) of 1071 (51 states multiplied by 21 years).

To construct this dataset, tables from NIEER and the US Census Bureau were merged based on state and year, creating a unified table of the relevant available information on this topic. The existence of early education programs, as defined by NIEER, is a key aspect of this dataset.

Quality benchmarks, crucial for assessing the effectiveness of these programs, are represented categorically, with "Yes" and "Program level only" indicating compliance with the criteria.

It is noteworthy that certain quality benchmarks were subject to changes. Consequently, these benchmarks cannot be considered throughout the entire duration of the dataset, except when counting the number of metrics met. In each year, there are 10 quality metrics. A summary of all quality metrics can be found in appendix A.

It is important to note that there are some instances in which certain variables were not reported. In such cases, the variable is interpolated from the years before and after. This estimation allows for a standardized approach to handling the few missing instances. Imputation, in this context, involves estimating missing variable values based on the available data points. The goal is to create a continuous representation of variable trends over the specified years. While this method introduces assumptions about the nature of variable changes, it allows for a more complete analysis of early childhood education programs.

Figure 1: Summary Statistics of Key Variables

Attribute	Avg	StDev	Min	Max	Count
Outcome: Poverty	12%	3%	4%	26%	1071
Program Indicators					
Program 3 Year Old Indicator	55%	50%	0%	100%	1071
Program 4 Year Old Indicator	82%	39%	0%	100%	1071
Spending					
All Spending per Pupil	$6,378	$4,548	$0	$24,405	1071
State Spending per Pupil	$5,265	$4,079	$0	$24,405	1071
Enrollment					
3yo Enrolled	4%	10%	0%	80%	1071
4yo Enrolled	21%	22%	0%	100%	1071
Quality Benchmarks *(Ref. Appendix A)*					
Quality Standards Met	43%	37%	0%	100%	1071
1. Early Learning & Development Standards	70%	46%	0%	100%	1071
3. Teacher Degree	51%	50%	0%	100%	1071
4. Teacher Specialized Training	70%	46%	0%	100%	1071
5. Assistant Teacher Degree	41%	49%	0%	100%	1071
6. Staff Professional Development	54%	50%	0%	100%	1071
7. Maximum Class Size	68%	47%	0%	100%	1071
8. Staff to Child Ratio	69%	46%	0%	100%	1071
9. Screening and Referral	60%	49%	0%	100%	1071
10. Continuous Quality Improvement Sys.	66%	47%	0%	100%	1071

Source: Poverty from US Census, Pre-K from NIEER.

Figure 1 shows summary statistics of key variables across all State-Year combinations.

All variables expressed in percentages where possible for maximum comparability.

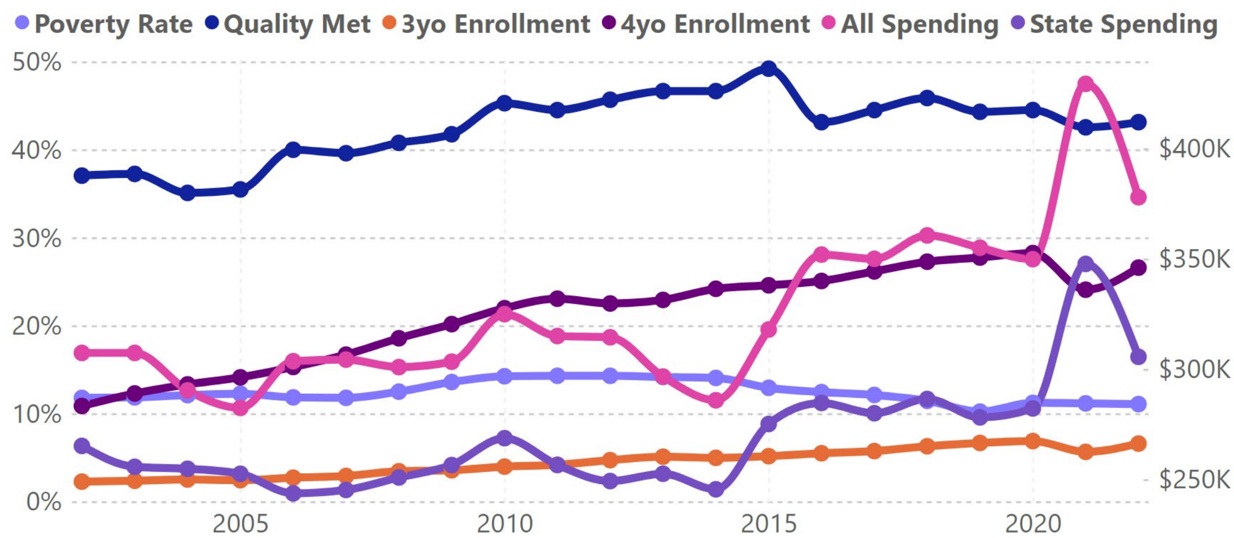
Notable observations include:

Programs for 3-year-olds are in their fledgling stages, as evidenced by average 4-year-old enrollment rates surpassing them by 17 percentage points.

On average, state spending per pupil is about $1.1k less than all spending per pupil, indicating most funding comes from the state level.

Across all states and years, 82% had a program for 4-year-olds, highlighting a limitation within the NIEER dataset – limited temporal scope.

Figure 2: Average Values of All Key Variables



Source: Poverty from US Census, Pre-K from NIEER.

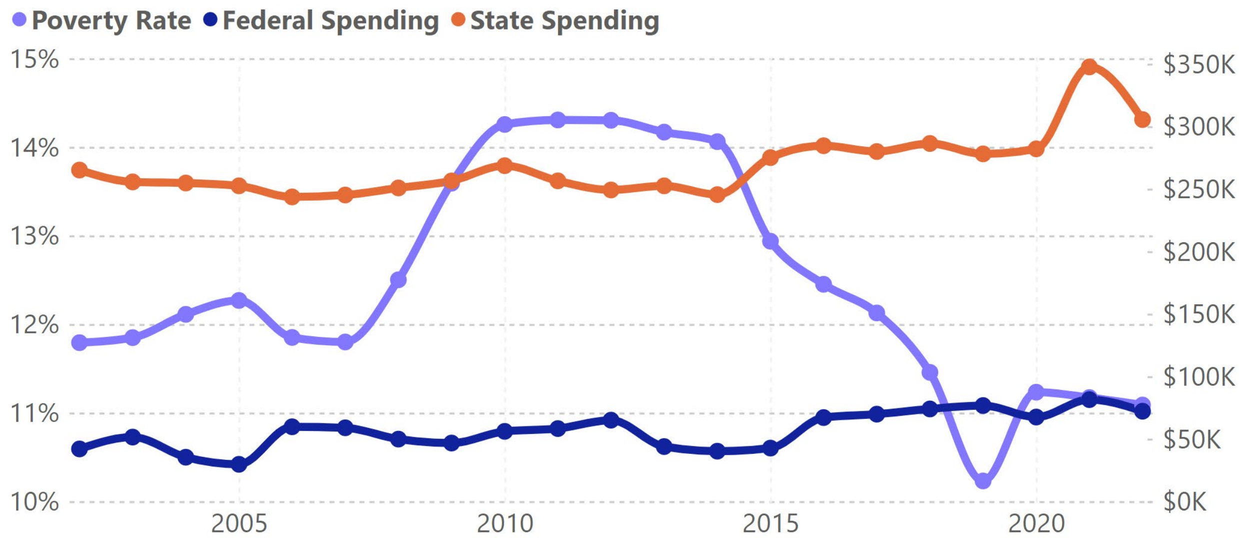
Figure 2 displays average values of key variables over time aggregated by Year. Due to the nature of the data, Per-Pupil spending is the rightmost Y-Axis, while all other variables, expressed in percentages, are on the leftmost Y-Axis. There does not appear to be any immediate major relationship between any of the independent variables and poverty rates, using cursory descriptive analytics.

Notable observations:

Enrollment and Poverty are ‘sticky’ variables, meaning they do not respond as much as other variables, such as spending or quality. This is particularly evident in 2021 with a large spike in spending.

When spending spikes in 2021, poverty rate experiences a small but notable dip, once again highlighting the stickiness, but also reinforcing a possible slight negative correlation. Admittedly, such an observation could be attributed to the Covid-19 pandemic after-effects.

Figure 3: Average Values of Spending and Poverty



Source: Poverty from US Census, Pre-K from NIEER.

Figure 3 highlights not only the relationship between spending and poverty, but also the interesting relationship between Federal and State Spending. On average, states provide most funding for their Pre-K programs, with federal funds accounting for about a third of total spending.

V. Theory

The conceptual framework of this research is rooted in economic theory, aiming to explore the multifaceted impact of state-funded Pre-K programs on poverty rates in the United States. The central economic theory guiding this study posits that subsidized Pre-K programs can influence poverty rates through two primary mechanisms: increasing parental workforce participation and alleviating the financial burden of childcare costs. Quality, spending, and enrollment are considered as modifiers that interact with these mechanisms, shaping the overall effectiveness of Pre-K programs in reducing poverty.

A. Increased Parental Workforce Participation:

The underlying assumption is that access to subsidized early education programs encourages parents, particularly mothers, to participate more actively in the workforce. By providing a structured and reliable childcare option, Pre-K programs can reduce the barriers faced by parents in seeking employment or increasing work hours.

B. Alleviation of Childcare Costs:

Accessible and high-quality Pre-K programs alleviate the financial burden of childcare costs on families, especially those with lower incomes. This reduction in childcare costs may free up financial resources that can be directed towards other essential needs, contributing to a decline in poverty rates.

C. Modifiers: Quality, Spending, and Enrollment

The quality of Pre-K programs is expected to play a crucial role in shaping their impact on poverty rates. High-quality programs with well-qualified teachers, effective curriculum design, and supportive infrastructure are anticipated to yield more positive outcomes for children, potentially enhancing the program's effectiveness in reducing poverty.

Adequate financial support is considered a foundational element for the success of Pre-K programs. States with higher per-pupil spending are expected to have the resources necessary to maintain program quality over time, ensuring sustained positive effects on poverty rates.

The accessibility and reach of Pre-K programs, as indicated by enrollment rates, are critical for understanding their overall effectiveness. Higher enrollment rates are expected to correlate with a more significant reduction in poverty rates, reflecting the broader societal impact of well-attended Pre-K programs.

VI. Methodology

This analysis rests on understanding the interactions between various mechanisms and modifiers that influence poverty rates. The state-year TWFE method is used to compare poverty rates across states over time. Temporal events such as economic crises are controlled via time fixed effects, while state fixed effects control for unique state-level characteristics like socioeconomic variables. This framework allows for isolating the effects of policy changes and economic factors, such as variations in program quality, funding, and enrollment rates, on poverty rates.

A. Primary model:

$$\text{Poverty Rate}_{st} = \beta_0 + \beta_1\text{Funding}_{st} + \beta_2\text{Enrollment}_{st} + \beta_3\text{Quality}_{st} + \text{State}_s + \text{Year}_t + \epsilon_{st} \quad (1)$$

B. Dependent Variable:

Poverty Rate: The poverty rate within a given state-year combination.

C. Independent Variables:

Spending: Total and State Per-pupil pre-k spending, measured in thousands of real 2022 U.S. dollars.

Enrollment: 3 and 4-year-old Enrollment rates for pre-k programs, expressed as a percentage.

Quality: Quality Standards Met, assessed on a scale of 1 to 10.

D. Model Controls:

State: State Fixed-Effects capturing persistent differences between states that could influence poverty rates.

Year: Time Fixed-Effects accounting for nationwide events or trends affecting poverty across all states in a particular year.

ϵ: The residual error for a given state-year.

Additional models focus on isolating each independent variable (funding, enrollment, quality) against the dependent variable (poverty rate) while retaining the TWFE framework.

E. Measurement Details:

Spending: Expressed in real 2022 U.S. dollars to ensure consistency in purchasing power.

Poverty and Enrollment Rates: Measured as percentages to reflect their proportional impact.

Quality Standards: Rated on a 10-point scale, capturing the comprehensive quality of pre-k programs.

VII. Results

Figure 4: Models

Regressor	Main	Spending	Enrollment	Quality
All Spending per Pupil (USD, Thousands)	0.07* (0.04)	0.07** (0.04)		
State Spending per Pupil (USD, Thousands)	-0.11** (0.05)	-0.11** (0.05)		
3 Year Old Enrollment Rate	-0.90 (1.41)		-0.93 (1.43)	
4 Year Old Enrollment Rate	0.14 (0.59)		0.05 (0.58)	
Quality Standards Met (N/10)	0.02 (0.02)			0.02 (0.02)
Intercept	8.63*** (0.37)	8.61*** (0.36)	8.63*** (0.36)	8.61*** (0.36)
Number of Observations	1071	1071	1071	1071
Adjusted R-Square	0.8359	0.8362	0.8354	0.8356
Overall Significance	75.54***	78.49***	78.73***	79.77***

*Source: Poverty by US Census, Pre-K Attributes by NIEER. Notes: *, **, and *** signify 10%, 5%, and 1% significance levels, respectively. Robust Standard Errors in parenthesis. State/Year Fixed Effects present in all Models.*

A. Interpretation of Coefficients

All models exhibit statistical significance at a 1% level and demonstrate strong explanatory power, with an Adjusted R-Square of around 0.83. The primary model shows that for every thousand dollars spent per student on a state-level pre-k program, poverty rates decline by 0.11 percentage points. In contrast, a similar increase in overall spending per pupil leads to a 0.07 percentage point increase in poverty rates. These coefficients are significant at the 5% and 10% levels, respectively. This pattern is corroborated in the isolated spending model, where each coefficient achieves statistical significance at the 5% level. The coefficients for spending are promising, since as demonstrated in Figure 3, most funding comes from the state level, rather than the federal level.

A 100-percentage point increase in enrollment rates for 3-year-olds leads to a 0.90 percentage point reduction in poverty rates. However, a similar increase in 4-year-old enrollment rates results in a 0.14 percentage point increase in poverty, but neither coefficient is statistically significant. In the isolated enrollment model, these coefficients are slightly more extreme but still statistically insignificant at standard levels.

Quality Standards Met, which is rated on a 10-point scale, shows that each additional quality standard met correlates with a small 0.02 percentage point increase in poverty rates. This coefficient remains consistent even in isolation.

Certain results deviate from the initial theory. Increased enrollment rates for 4-year-olds, total spending per pupil, and meeting quality benchmarks all have positive coefficients. These relationships remain similar when each factor is analyzed independently.

Despite the relatively small coefficients, my findings suggest that Pre-K programs have a place in poverty alleviation, though it is necessary to consider the limitations of the broad approach this research uses without having appropriate control variables to limit the scope to the groups subsidized Pre-K targets. The coefficients, removing appropriate context, are not economically significant, which as previously stated can be attributed to overly broad analysis without necessary control variables to add precision.

B. Reflection

Overall, my theory posits that subsidized pre-k programs impact a relatively narrow demographic subset: parents of 3- and 4-year-old children. This leads to considerable noise in the analysis due to the approach not accounting for factors like age and gender. Unfortunately, the scope of the project, limited by time and computational constraints, prevents a more targeted methodology, as the available data is challenging to handle and requires careful planning for execution and processing time.

VIII. Conclusion

This study demonstrates that state-funded Pre-K programs can have a meaningful impact on poverty rates in the United States. While the analysis used state-year fixed effects to reveal relationships between spending, enrollment, and quality indicators with poverty outcomes, future research could refine our understanding of the mechanisms involved by using more detailed data.

Firstly, the signs of the coefficients align with expectations when compared to their complementary variables, such as enrollment rates for 3- and 4-year-olds, or state versus total spending. The coefficients are also appropriately sized, considering they affect a relatively narrow demographic subset, particularly parents of 3- and 4-year-old children. Future research should consider incorporating individual-level data to add more controls, such as age, gender, and parental workforce participation. Despite computational and time constraints that limited this study's scope, such data would improve precision and offer a clearer picture of the programs' impact.

Secondly, introducing more localized geographical fixed effects, like county or district levels, could provide a detailed understanding of the relationship between school funding, enrollment, and quality. While this level of detail is currently unattainable due to public data limitations, it would be valuable in identifying regional disparities in program effectiveness and ensuring equitable policy improvements.

Lastly, this paper highlighted that increased state funding per child correlates with a reduction in poverty rates, underscoring the importance of sustained and equitable Pre-K program financing. More comprehensive methodologies would also help explain unexpected results, such as positive correlations between poverty rates and quality benchmarks met or 4-year-old enrollment. This deeper exploration would enable policymakers to refine interventions and optimize early childhood education policies.

Overall, while this paper showed promising correlations between Pre-K program characteristics and poverty outcomes, future research should aim for greater precision in addition to geographical coverage to guide impactful educational policies.

IX. References:

Barnett, W. S. (2003). Better Teachers, Better Preschools: Student Achievement Linked to Teacher Qualifications. NIEER Preschool Policy Matters, Issue 2.

Bartik, T. J., & Hershbein, B. J. (2018). Pre-K in the public schools: Evidence from within US states (No. 18-285). Upjohn Institute working paper.

Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. Applied developmental science, 6(1), 42-57.

National Institute for Early Education Research. [dataset] (2023). https://nieer.org/data-explorer Data extracted on April 22, 2024

Friedman-Krauss, A. H., Barnett, W. S., Garver, K. A., Hodges, K. S., Weisenfeld, G. G., Gardiner, B. A., & Jost, T. M. (2023). The State of Preschool 2021: State Preschool Yearbook. National Institute for Early Education Research.

Lamy, C. E. (2013). How preschool fights poverty. Educational Leadership, 70(8), 32-36.

Manning, M., & Patterson, J. (2006). LIFETIME EFFECTS: The High/Scope Perry preschool study through age 40. Childhood Education, 83(2), 121.

Martinez, S., Naudeau, S., & Pereira, V. A. (2017). Preschool and child development under extreme poverty: evidence from a randomized experiment in rural Mozambique. World Bank Policy Research Working Paper, (8290).

McCoy, D. C., Connors, M. C., Morris, P. A., Yoshikawa, H., & Friedman-Krauss, A. H. (2015). Neighborhood economic disadvantage and children's cognitive and social-emotional development: Exploring Head Start classroom quality as a mediating mechanism. Early childhood research quarterly, 32, 150-159.

Pearman, F. A. (2020). The moderating effect of neighborhood poverty on preschool effectiveness: Evidence from the Tennessee voluntary prekindergarten experiment. American Educational Research Journal, 57(3), 1323-1357.

Reynolds, A. J. (1994). Effects of a preschool plus follow-on intervention for children at risk. Developmental Psychology, 30(6), 787.

Reynolds, A. J., Ou, S. R., Mondi, C. F., & Giovanelli, A. (2019). Reducing poverty and inequality through preschool-to-third-grade prevention services. American Psychologist, 74(6), 653.

Rossin-Slater, M., & Wüst, M. (2020). What is the added value of preschool for poor children? Long-term and intergenerational impacts and interactions with an infant health intervention. American Economic Journal: Applied Economics, 12(3), 255-286.

Schippers, V. (2014). No Preschooler Left Behind: The Need for High Quality Early Intervention for Children Born into Poverty. Multicultural Education, 22(1), 41-45.

Schroeder, J. (2007). Full-day kindergarten offsets negative effects of poverty on state tests. European Early Childhood Education Research Journal, 15(3), 427-439.

Schweinhart, L. J., Barnes, H. V., & Weikhart, D. P. (2005). Significant benefits: The High/Scope Perry preschool study through age 27. Child welfare: Major themes in health and social welfare, 4, 9-29.

US Census Bureau. percent of persons in poverty, by state - census.gov. [dataset] (n.d.). https://www2.census.gov/programs-surveys/cps/tables/time-series/historical-poverty-people/hstpov19.xls Data extracted on April 22, 2024

Valentino, R. (2018). Will public Pre-K really close achievement gaps? Gaps in prekindergarten quality between students and across states. American Educational Research Journal, 55(1), 79-116.

X. Appendices

Appendix A: Definitions of Quality Benchmarks (NIEER, 2023)

1. **Early Learning and Development Standards (ELDS):** Requires comprehensive and specific ELDS, aligned with state standards and child assessments, supported by professional development for preschool-aged children.
2. **Curriculum Supports (Added 2016):** Requires states to offer guidance or approval processes for selecting and implementing curricula, ensuring strong support for learning and development in language, literacy, mathematics, and social-emotional domains.
3. **Teacher Degree:** Requires lead teachers in every classroom to have at least a bachelor's degree, aligning with research advocating for well-qualified teachers to provide high-quality educational environments.
4. **Teacher Specialized Training:** Requires policies to mandate specialized training in early childhood education and/or child development for preschool lead teachers, recognizing the importance of tailored preparation.
5. **Assistant Teacher Degree:** Requires policies to mandate that assistant teachers hold a Child Development Associate (CDA) credential or equivalent preparation based on coursework, emphasizing the benefits of preservice preparation.
6. **Staff Professional Development:** Requires both teachers and assistant teachers to have at least 15 hours of annual in-service training, including coaching or classroom-embedded support, with individualized professional development plans for continuous improvement.
7. **Maximum Class Size:** Requires a class size of no greater than 20 children.
8. **Staff-Child Ratio:** Requires classes to have no more than 10 children per classroom teaching staff member.

9. **Screenings and Referrals:** Requires preschool programs to ensure vision, hearing, and additional health screenings for children, with referrals as needed, recognizing the importance of overall well-being and educational success.
10. **Continuous Quality Improvement System (CQIS):** Requires an effective CQIS, involving the systematic collection of data on classroom quality, with both local programs and the state using this information for continuous improvement, emphasizing a cycle of planning, observation, and feedback.
11. **Meals (Discontinued, 2002-2015):** Requires meals to be served with the program. NIEER states this was unintentionally an indicator of if a program was full or half day and has therefore been discontinued.

Appendix B: Python Dataset Creation Codes

```
#################################################################
### Full Repo Available at: https://github.com/Nathan-D-R/UPK ###
#################################################################

## Libraries
import pandas as pd
import requests
import os
import subprocess

def download_file(url, directory="Data", filename=None):
    if filename is None:
        filename = url.split('/')[-1]
    os.makedirs(directory, exist_ok=True)
    filepath = os.path.join(directory, filename)
    response = requests.get(url)
    response.raise_for_status()
    with open(filepath, 'wb') as f:
        f.write(response.content)
    print(f"Downloaded {filename} to {filepath}")

def run_python(script_name):
    subprocess.run(["python", script_name], check=True)
    print(f"Ran script {script_name}")

def run_sas(script_name):
    subprocess.run(["sas", script_name], check=True)
    print(f"Ran script {script_name}")

urls = [
    "https://www2.census.gov/programs-surveys/cps/tables/time-
series/historical-poverty-people/hstpov19.xlsx",
```

```
    "https://nieer.org/sites/default/files/2023-
12/state_preschool_quality_standards_met.xlsx",
    "https://nieer.org/sites/default/files/2023-
12/state_preschool_spending.xlsx",
    "https://nieer.org/sites/default/files/2023-
12/state_preschool_enrollment.xlsx"
]


## Data Cleaning Functions


def clean_poverty(file):
    # import Data/hstpov19.xlsx
    poverty = pd.read_excel(file, sheet_name='pov19', skiprows=3,
usecols='A:F', header=None)


    # New column '6', copy value from 1 if it starts with '20'
    poverty[6] = poverty[0].apply(lambda x: x if str(x).startswith('20') else
None)


    # Fill down column '6'
    poverty[6] = poverty[6].fillna(method='ffill')


    # Filter where 1 is not null or is 'Total'
    poverty = poverty[poverty[1].notnull()]
    poverty = poverty[poverty[1] != 'Total']


    # Promote headers from row 0
    poverty.columns = poverty.iloc[0]



    # Rename Columns
    poverty.columns = ['State Name', 'Total', 'Number', 'N SE', 'Poverty', 'P
SE', 'Year']


    # Reorder columns
```

```
    poverty = poverty[['State Name', 'Year', 'Total', 'Number', 'N SE',
'Poverty', 'P SE']]

    poverty = poverty[poverty['State Name'] != 'State']

    # Print unique values in 'Year'
    print(poverty['Year'].unique())

    # Replace values in 'Year' column
    poverty['Year'] = poverty['Year'].replace(
        {'2020 (1)': '2020', '2013 (3)': '2013', '2010 (5)': '2010', '2004
(6)': '2004'}
    )

    # Drop where year contains '('
    poverty = poverty[~poverty['Year'].str.contains('\(')]

    # Year as int
    poverty['Year'] = poverty['Year'].astype(int)

    # Filter where 'Year' >= 2002
    poverty = poverty[poverty['Year'] >= 2002]

    # Drop all but State Year and Percent
    poverty = poverty[['State Name', 'Year', 'Poverty']]

    return poverty

def clean_quality(file):
    # Read in data
    quality = pd.read_excel(file)

    # Filter where 'Program Name' = null
    quality = quality[quality['Program Name'].isnull()]
```

```
    # Drop 'Program Name' column
    quality = quality.drop(columns=['Program Name'])

    # Replace values in 'Variable Name' column
    quality['Variable Name'] = quality['Variable Name'].replace({
        'Family Support Service Requirements Benchmark': 'Continuous Quality
Improvement System Benchmark',
        'Monitoring Benchmark': 'Continuous Quality Improvement System
Benchmark',
        'Early Learning Standards Benchmark': 'Early Learning & Development
Standards Benchmark',
        'Teacher In-Service Benchmark': 'Staff Professional Development
Benchmark'
    })

    # Pivot 'Variable Name' column
    quality = quality.pivot_table(index=['State Name', 'Year'],
columns=['Variable Name'], values='Value', aggfunc='first').reset_index()

    # Add new column to count number of "Yes" (not 'State Name' or 'Year'')
    quality['Quality Standards Met'] = quality.drop(columns = ['State Name',
'Year']).apply(lambda x: x.str.contains('yes', case=False).sum(), axis=1)

    return quality

def clean_general(file):
    # Read in data
    data = pd.read_excel(file)

    # Filter where 'Program Name' = null
    data = data[data['Program Name'].isnull()]

    # Drop 'Program Name' column
    data = data.drop(columns=['Program Name'])
```

```
    data = data.rename(columns={'Spending (2022 Dollars)': 'Value'})


    data = data.rename(columns={'Enrollment': 'Value'})



    # Pivot on 'Variable Name'
    data = data.pivot_table(index=['State Name', 'Year'], columns=['Variable
Name'], values='Value', aggfunc='first').reset_index()


    return data


# Merge Data by 'State Name' and 'Year'
def merge_data(poverty, spending, enrollment, quality):
    # Merge poverty and spending
    data = pd.merge(poverty, spending, on=['State Name', 'Year'], how='outer')
    # Merge data and enrollment
    data = pd.merge(data, enrollment, on=['State Name', 'Year'], how='outer')
    # Merge data and quality
    data = pd.merge(data, quality, on=['State Name', 'Year'], how='outer')


    # Replace values in all columns
    data = data.replace({'Yes': 1, 'No': 0, 'No program': 0, 'NA - Program
level only': 1, '': 'NOT COLLECTED', 'Not reported': 'NOT REPORTED'})


    # Program Indicators
    data['Program_3yo'] = data['Percentage of 3-year-olds Enrolled in State
Pre-K'].apply(lambda x: 0 if x == 0 else 1)


    data['Program_4yo'] = data['Percentage of 4-year-olds Enrolled in State
Pre-K'].apply(lambda x: 0 if x == 0 else 1)


    # Sort by 'State Name' and 'Year'
    data = data.sort_values(by=['State Name', 'Year']).reset_index(drop=True)
```

```
    return data


def rename_columns(data):
    data = data.rename(columns={'State Name': 'State'})
    data = data.rename(columns={'Poverty Rate': 'Poverty'})
    data.columns = data.columns.str.replace(' ', '_')
    data.columns = data.columns.str.replace('_\(2022_Dollars\)', '',
regex=True)
    data.columns = data.columns.str.replace('All-Reported', 'All')
    data.columns = data.columns.str.replace('Total_', '')
    data.columns = data.columns.str.replace('_in_State_Pre-K', '')
    data.columns = data.columns.str.replace('_Benchmark', '_B')
    data.columns = data.columns.str.replace('3-year-olds', '3yo')
    data.columns = data.columns.str.replace('4-year-olds', '4yo')
    data.columns = data.columns.str.replace('Percentage_of', 'P')
    data.columns = data.columns.str.replace('Number_of', 'N')
    data.columns = data.columns.str.replace('&', 'and')



    return data


def fill_missing(data):
    # List of columns to apply the filling logic
    numeric_cols = [
        'All_Spending_per_Child', 'State_Spending_per_Child',
        'All_Spending', 'State_Pre-K_Spending', 'N_3yo_Enrolled',
        'N_4yo_Enrolled', 'P_3yo_Enrolled', 'P_4yo_Enrolled',
        'State_Pre-K_Enrollment', 'Assistant_Teacher_Degree_B',
        'Continuous_Quality_Improvement_System_B',
        'Early_Learning_and_Development_Standards_B',
        'Maximum_Class_Size_B', 'Screening_and_Referral_B',
        'Staff_Professional_Development_B',
        'Staff_to_Child_Ratio_B', 'Teacher_Degree_B',
```

```
        'Teacher_Specialized_Training_B',
        'Quality_Standards_Met'
    ]

    # Replace "NOT REPORTED" with NaN (null) in the specified columns
    data[numeric_cols] = data[numeric_cols].replace("NOT REPORTED", pd.NA)

    # Convert columns with potential textual representations of numbers to
numeric, forcing errors to NA
    for col in numeric_cols:
        data[col] = pd.to_numeric(data[col], errors='coerce')

    # Group by state to ensure that the filling logic is applied within each
state
    grouped = data.groupby('State')

    # Function to apply the specified filling logic to a column within each
group
    def fill_na_within_group(series):
        # Forward fill then backward fill for ends
        series = series.fillna(method='ffill').fillna(method='bfill')
        # Interpolate for mid values
        return series.interpolate()

    # Apply the filling logic to each group for the specified numeric columns
    for col in numeric_cols:
        data[col] = grouped[col].transform(fill_na_within_group)

    return data

def main():
    # Download Source Files
    for url in urls:
        download_file(url)
```

```python
    # Import and clean data
    poverty = clean_poverty('./Data/hstpov19.xlsx')
    quality = clean_quality('./Data/state_preschool_quality.xlsx')
    spending = clean_general('./Data/state_preschool_spending.xlsx')
    enrollment = clean_general('./Data/state_preschool_enrollment.xlsx')

    # Merge data on 'State Name' and 'Year'
    data = merge_data(poverty, spending, enrollment, quality)

    data = rename_columns(data)

    data = fill_missing(data)

    # Remove States "National" and "Guam"
    data = data[data['State'] != 'National']
    data = data[data['State'] != 'Guam']

    # Export to data.xlsx
    data.to_excel('./data.xlsx', index=False)

# Entry Point
if __name__ == '__main__':
    main()
```

Appendix C: SAS Regression Codes

```
/*********************************************************/
/* Full code available at https://github.com/Nathan-D-R/UPK */
/*********************************************************/


/* Define the function */
%macro CleanOutput(input);


data Results;
      set PE;
      where SUBSTR(Parameter, 1, 6) ne "State " and SUBSTR(Parameter, 1, 4) ne
"Year";
run;


data Results;
      length Model $5;
      length Parameter $30;
      set Results;
      Model = "Model";
      if Probt le 0.01 then Star="***";
            else if Probt le 0.05 then Star="**";
            else if Probt le 0.1 then Star="*";

            /* Handle stars */
      EditedResults=cats(Put(Estimate,comma16.2),star);
      output;

      /* Handle robust standard errors */
      EditedResults=cats("(",put(StdErr,comma16.2),")");
      output;
run;


data Results;
```

```
    set Results;
    if mod(_n_,2)=0 then Parameter = "x-" || trim(Parameter);
run;

data Results;
    set Results;
    Regressor = Parameter;
    Result = EditedResults;
    keep Regressor Result;
run;

data NumofObs;
    set OBS(rename=() drop=CValue1);
    where Label1="Number of Observations";
    Model=put(nValue1,comma16.);
    drop nValue1;
run;

data AdjRsq;
    set AdjRsq(rename=(cvalue1=Model) drop=nvalue1);
    Where Label1 = "Adjusted R-Square";
run;

data OSM;
    set OverallSig;
    where Effect="Model";
    if ProbF le 0.01 then Star="***";
        else if ProbF le 0.05 then Star="**";
        else if ProbF le 0.1 then Star="*";

    Label1="Overall Significance";
    EditedValue=cats(put(FValue,comma16.2),Star);
    Model = EditedValue;
```

```
	keep Label1 Model;
run;

/* Combine rows for the other statistics */
data OtherStat;
	length Model $32;
	set NumofObs AdjRsq OSM;
	rename Label1=Regressor Model=Result;
run;

data Model&input.;
	set Results OtherStat;
run;

data Model&input.;
	set Model&input.(rename=(Result=Model&input.));
	retain Index&input. 0;
	Index&input. + 1;
run;

proc sort data=Model&input.;
	by Regressor;
run;

%mend;

*%include "/home/u62949701/MySAS/CleanOutput.sas";

%macro GenerateModels();
%let n = 1;

proc import datafile="/home/u62949701/MySAS/Data/data.xlsx"
```

```
	out=work.Data
	dbms=xlsx
	replace;
run;

data Data;
	set Data;
	if State = "District of Columbia" then delete;
run;

/* All Variables */
ods output ParameterEstimates=PE DataSummary=Obs
		  FitStatistics=AdjRsq Effects=OverallSig;
proc surveyreg Data=Data;
	Class State Year;
	Model Poverty =
	All_Spending_per_Child
	State_Spending_per_Child
	P_3yo_Enrolled
	P_4yo_Enrolled
	Quality_Standards_Met
	State Year /Solution AdjRsq;
run;

%CleanOutput(&n); %let n = %eval(&n + 1);

/* Main */
ods output ParameterEstimates=PE DataSummary=Obs
		  FitStatistics=AdjRsq Effects=OverallSig;
proc surveyreg Data=Data;
	Class State Year;
	Model Poverty =
	All_Spending_per_Child
```

```
      State_Spending_per_Child
      P_3yo_Enrolled
      P_4yo_Enrolled
      Quality_Standards_Met
      State Year /Solution AdjRsq;
run;

%CleanOutput(&n); %let n = %eval(&n + 1);

/* Spending */
ods output ParameterEstimates=PE DataSummary=Obs
                  FitStatistics=AdjRsq Effects=OverallSig;
proc surveyreg Data=Data;
      Class State Year;
      Model Poverty =
      All_Spending_per_Child
      State_Spending_per_Child
      State Year /Solution AdjRsq;
run;

%CleanOutput(&n); %let n = %eval(&n + 1);

/* Enrollment */
ods output ParameterEstimates=PE DataSummary=Obs
                  FitStatistics=AdjRsq Effects=OverallSig;
proc surveyreg Data=Data;
      Class State Year;
      Model Poverty =
      P_3yo_Enrolled
      P_4yo_Enrolled
      State Year /Solution AdjRsq;
run;
```

```
%CleanOutput(&n); %let n = %eval(&n + 1);

/* Quality */
ods output ParameterEstimates=PE DataSummary=Obs
                  FitStatistics=AdjRsq Effects=OverallSig;
proc surveyreg Data=Data;
      Class State Year;
      Model Poverty =
      Quality_Standards_Met
      State Year /Solution AdjRsq;
run;

%CleanOutput(&n); %let n = %eval(&n + 1);


%let n = %eval(&n - 1);

data Final;
      merge Model1-Model&n;
      by Regressor;
run;

proc sort data=Final;
      by Index1;
run;

data Final;
      set Final;
      keep Regressor Model2-Model&n;
      if SUBSTR(Regressor, 1, 2) = "x-" then Regressor = "";
run;
```

```
ods excel file="/home/u62949701/MySAS/Exports/SeniorProject.xlsx"
options(Embedded_Titles="ON" Embedded_Footnotes="ON");

proc print data=Final noobs;

      var Regressor;

      *format Regressor $Regressor.;

      var Model2-Model&n /

      style(header)={Just=Center}

      style(data)={Just=Center tagattr="type:string"};

      title "Figure: Models";

      footnote "Source: Poverty by US Census Bureau, Pre-K Attributes by
NIEER.";

      footnote2 "Notes: *, **, and *** signify 10%, 5%, and 1% significance
levels, respectively.";

      footnote3 "Robust Standard Errors in parenthesis.";

run;

ods excel close;


%mend;

%GenerateModels();
```