Title:
A Reanalysis of the Impacts of the Tennessee Voluntary Prekindergarten Program

Authors:
Tyler W. Watts (New York University), Greg J. Duncan (University of California, Irvine), Mariela Rivas (University of California, Irvine)

Background/Context:
The evidence base on preschool programs is large and growing, but we severely lack evaluations of preschool programs that use experimental designs. Most of the current research reporting long-run preschool effects has depended on non-experimental or quasi-experimental designs, leaving questions as to whether the widespread scale-up of public preschool should be expected to produce long-lasting effects on children's lives.

The Tennessee Voluntary Prekindergarten Program (TNVPK) was recently evaluated with a lottery-based random assignment design (Lipsey, Farran, & Durkin, 2018). However, study evaluators made a number of analytic decisions that were not typical of the standard program evaluation techniques found in the economics literature. In particular, Lipsey and colleagues (2018) defined the "treatment group" in a way that adjusted for enrollment in non-study sites, opening the door for potential selection bias. Further, although random assignment was conducted at the site level, Lipsey and colleagues did not control for site fixed effects, raising questions as to whether reported effects could be due to between site differences that should be controlled.

Purpose/Objective/Research Question:
We provide an independent re-analysis of TNVPK data, relying on econometric approaches to estimating the effects of TNVPK. We examine the ITT impacts of the opportunity to enroll in TNVPK on a host of child outcomes measured through the end of third grade, and we also address issues surrounding non-compliance, attrition, site heterogeneity and fadeout through a number of additional analyses.

Setting:
The TNVPK evaluation relied on participation from over-subscribed pre-k sites, which agreed to implement a lottery to determine student enrollment (see below). These sites were drawn from urban, suburban and rural areas.

Population/Participants/Subjects:
The TNVPK program is targeted at low-income 4-year-olds, and almost all children participating in the study were eligible for Free or Reduced Price Lunch. Approximately 48% identified as White, 27% as Black and 23% as Hispanic.

Intervention/Program/Practice:
The TNVPK program was introduced by the state to increase the school readiness skills of children from low-income families. In 2015, TNVPK met 9 of the 10 quality benchmarks set by the National Institute for Early Education Research (see http://nieer.org/state-preschool-yearbooks/yearbook2015).

Research Design:
The TN Department of Education (DOE) recruited pre-k sites that were regularly over-subscribed to implement a lottery procedure to determine enrollment. Participating sites (n=79) sent enrollment lists to the Peabody Research Institute (PRI) at Vanderbilt University, and PRI researchers randomly ordered the enrollment lists before returning lists to participating sites. Participating sites then used the randomized lists to offer TNVPK enrollment.

PRI researchers used enrollment data to determine who had been offered a chance to attend. Thus, students who appeared higher on a given list were considered part of the "treatment group" and students who were placed on the waitlist were considered part of the "control group." These lists were then matched with administrative data made available by the TN DOE, and students were tracked from kindergarten through the end of third grade using administrative information such as test scores, attendance records, records of disciplinary offenses, grade retention/promotion indicators and special education/ gifted and talented placements.

Data Collection and Analysis:
Tests of baseline balance indicated that the lottery process produced equivalent groups on observables. Our ITT estimates depend on modeling each outcome measure listed above as a function of "treatment status" and random-assignment-list fixed effects (i.e., we control for each respective random assignment list used to determine enrollment). We also include baseline demographic characteristics to improve precision. We supplement these ITT estimates with two-stage-least-squares models (2SLS) that attempt to recover the effect of actually *attending* TNVPK on later outcomes. Because descriptive information indicated substantial non-compliance in the control group (i.e., 89% of the treatment group attended TNVPK, while 41% of the control group also attended), the 2SLS are crucial for understanding the effect of participation in TNVPK on child educational outcomes.

Findings/Results:
Table 1 presents ITT estimates for our key outcomes taken from administrative data. In Columns 1 through 8, we present estimates for each respective academic year between kindergarten in grade 3. In Column 9, we present estimates for outcome measures that aggregated data cross the individual follow-up waves (e.g., "was a given child ever placed in special education between kindergarten and grade 3?"). For binary outcome measures (special education placement, gifted placement, serious disciplinary offenses, and retention), we estimated linear probability models. For continuous outcomes (test scores and absences), outcome variables were standardized and coefficients can be interpreted as effect sizes.

As Table 1 reflects, we found mainly null effects of an offer to attend TNVPK on gifted placements, disciplinary offenses, absences and retention. However, models suggested that TNVPK led to higher rates of placement in special education between kindergarten and grade 3 (4.4% higher rate across this period, $p < 0.05$), and slightly lower third grade test scores ($\beta = -0.081$, $p < 0.10$).

Table 2 presents effects for our 2SLS models, which estimated the impact of actually attending TNVPK on later outcomes. These 2SLS models suggests that TNVPK participation led to an increase in special education placement of 11.4 percentage points ($p < 0.05$) between kindergarten

and third grade, and the negative test score impact was also stronger in magnitude (β = - 0.207, $p <$ 0.10).

Conclusion:
The full paper presents results from models that adjusted for attrition, and we found no indication that attrition biased the estimates shown in Tables 1 or 2. We also explored patterns of site heterogeneity, and we found that site effects were normally distributed around the average effects reported in Table 1. The proposed presentation would also present further analyses attempting to unpack why TNVPK appeared to have a detrimental effect on later test scores, as we show that the program boosted cognitive skills measured at the end of pre-k.

In sum, despite employing different analytic approaches, our findings largely confirm the results reported by Lipsey and colleagues (2018) and suggest that more careful evaluation work is needed if pre-k programs implemented at scare are expected to produce positive long-term outcomes.

References
Lipsey, M. W., Farran, D. C., & Durkin, K. (2018). Effects of the Tennessee Prekindergarten Program on children's achievement and behavior through third grade. *Early Childhood Research Quarterly, 4,* 155-176.

Table 1

Impact Estimates for the TNVPK Program

	Kindergarten		Grade 1		Grade 2		Grade 3		Outcome Summation
	FE Only	Full Controls	FE Only	Full Controls	FE Only	Full Controls	FE Only	Full Controls	Full Controls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Special Ed	0.030**	0.032**	0.027+	0.030*	0.027	0.030+	0.019	0.021	0.044*
	(0.011)	(0.011)	(0.015)	(0.015)	(0.017)	(0.017)	(0.018)	(0.018)	(0.018)
CTL Mean	0.074		0.095		0.108		0.099		0.135
Gifted	-0.000	-0.000	-0.004	-0.004	-0.008	-0.008	-0.010	-0.009	-0.009
	(0.002)	(0.002)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.006)
CTL Mean	0.003		0.006		0.011		0.013		0.014
Discipline Off.	0.004	0.004	0.002	0.002	-0.010	-0.010	0.016+	0.016+	0.004
	(0.005)	(0.004)	(0.007)	(0.006)	(0.008)	(0.008)	(0.009)	(0.009)	(0.012)
CTL Mean	0.010		0.021		0.027		0.027		0.065
Absences (std)	0.074+	0.083+	0.022	0.029	0.040	0.049	0.046	0.047	0.056
	(0.044)	(0.044)	(0.043)	(0.043)	(0.039)	(0.039)	(0.042)	(0.042)	(0.041)
CTL Mean	0.002		0.000		0.002		-0.019		0.000
Retention	-0.015	-0.015	0.013	0.014	-0.006	-0.006	0.006	0.006	-0.001
	(0.010)	(0.010)	(0.009)	(0.009)	(0.005)	(0.005)	(0.004)	(0.004)	(0.013)
CTL Mean	0.052		0.029		0.014		0.005		0.099
Test Composite (std)	-	-	-	-	-	-	-0.089+	-0.081+	-
							(0.047)	(0.045)	
CTL Mean							0.000		
Controls									
R-List F.E.	Inc.	Inc.	Inc.	Inc.	Inc.	Inc.	Inc.	Inc.	Inc.
State Controls		Inc.		Inc.		Inc.		Inc.	Inc.
Observations	2876		2828		2783		2417		2925

Note. Robust standard errors were adjusted for site-level clustering and are presented in parentheses. Continuous variables (i.e., test scores and absences) were standardized using the control group mean and standard deviation for each wave, so coefficients can be likened to effect sizes. All other outcomes were binary, and estimates were generated from linear probability models. All models included fixed effects for the 111 TNVPK random assignment lists (r-list). Each set of estimates was derived from a separate model. For each respective outcome at each respective wave, we present two estimates. The first (i.e., the estimates shown in odd numbered columns) was derived from a model that only included the treatment status indicator and r-list fixed effects. The second (i.e., the estimates shown in even numbered columns) included the treatment status indicator, r-list fixed effects, and baseline demographic characteristics. The "outcome summation" column presents results indicating whether the outcome ever occurred over the measurement period for special education placements, gifted and talented section, serious disciplinary offenses, and retention. For absences, the "outcome summation" column models used the total number of absences recorded between kindergarten and grade 3. For each outcome, we also present the mean for the control group.

+ $p<0.10$ * $p< 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 2

TNVPK: Impacts Adjusted for Non-Compliance

	ITT	OLS	2SLS
	(1)	(2)	(3)
3rd Grade Test Composite	-0.081+	0.014	-0.207+
	(0.045)	(0.049)	(0.114)
Observations		2417	
Outcome Summations			
Special Education	0.044*	0.048**	0.114*
	(0.018)	(0.017)	(0.046)
Gifted and Talented	-0.009	-0.001	-0.023
	(0.006)	(0.005)	(0.015)
Behavioral Off.	0.004	-0.008	0.011
	(0.012)	(0.011)	(0.029)
Absences	0.056	-0.016	0.144
	(0.041)	(0.047)	(0.106)
Retention	-0.001	-0.026*	-0.003
	(0.013)	(0.013)	(0.033)
Observations		2925	

Note. Robust standard errors were adjusted for site-level clustering and are presented in parentheses. For the administrative data outcomes, we used the "outcome summation" variables (see Table 1). All models included random assignment list fixed effects and baseline controls. The ITT estimates are identical to the estimates shown in Table 1. The OLS estimates came from a series of regressions in which each outcome was regressed on a dummy variable for "whether attended 20 days" and covariates. The IV estimates came from a series of 2SLS runs, and variation in "whether attended 20 days" was produced solely by the random assignment indicator (i.e., the instrument).

+ p<0.10 * p< 0.05 ** p < 0.01 *** p < 0.001