Does One More Year Matter? Dosage Effect of the One-Village-One-Preschool Intervention in Rural China

Si Chen, Chen Zhao, Chen Chen, Zhiping Wu, Catherine E. Snow & Mai Lu

Routledge
Taylor & Francis Group

INTERVENTION, EVALUATION, AND POLICY STUDIES   OPEN ACCESS   Check for updates

Does One More Year Matter? Dosage Effect of the One-Village-One-Preschool Intervention in Rural China

Si Chen[a] (iD), Chen Zhao[b], Chen Chen[a], Zhiping Wu[b], Catherine E. Snow[a], and Mai Lu[b]

[a]Harvard Graduate School of Education, Cambridge, Massachusetts, USA; [b]China Development Research Foundation, Beijing, China

ABSTRACT

The One-Village-One-Preschool (OVOP) initiative aims to guarantee early childhood education (ECE) to all children in high-poverty villages in China. A challenge to policymakers is to balance expanding the scale with lengthening service duration. Following 23,775 children from preschool (4-year-old) to fourth-grade (10-year-old) in a poverty-stricken county, we found: 3-year and 2-year-OVOP groups started off the first-grade with a similar level of performance in Chinese, English, and math, lagging behind the well-resourced township-public-ECE group by 0.14 SDs. However, by the third or fourth-grade, the 3-year-OVOP group had emulated the township-public-ECE group, whereas the 2-year-OVOP group lagged by about 0.2 SDs. Our finding suggested that 1 more year in ECE for children in high-poverty-villages can counteract the long-term fade-out effect.

Early childhood education (ECE) is, in the long run, the most effective and cost-efficient intervention to leverage when addressing the inequalities that occur due to poverty (Engle et al., 2011). Great efforts have been made worldwide to grant ECE accessibility to low-income communities. As initial access has been widely established, scholars and practitioners have started looking into the necessity for improved length of ECE service for individuals in poverty (Jenkins et al., 2016). Would more years of ECE pay off with improved long-term development? This question, which is of interest to all policy makers who operate within the constraints of limited budgets, is of preeminent importance in the context of China: in 2009, the China Development Research Foundation (CDRF) of the State Council of China launched the One Village One Preschool (OVOP) initiative, which aims to guarantee ECE access to all children in China. This initiative will be written into national policy in the next 3 years. Promising levels of progress have been made since the initiative's launch; however, what remains to be decided is if ECE service

should cover 2 years (equivalent to 1 year of pre-kindergarten (PreK) + 1 year of kinder-garten in the U.S. context) or 3 years (equivalent to 2 years of PreK + 1 year of kinder-garten) of services for each child (To clarify, the 3-year OVOP group started at age 4, and the 2-year OVOP group started at age 5). This study aimed to address this question by comparing the long-term achievement (up to 4th grade in elementary school) of children in high poverty communities in rural China who received 2 or 3 years of the OVOP intervention compared to those who received early childhood services in other auspices or not at all.

Background

Early Childhood Intervention for Low-Income Families

Poverty poses social, psychological, biological and safety threats to child development (Walker et al., 2011). Researchers have pointed to ECE as the most effective leverage against poverty-induced inequality in social and human capital (Engle et al., 2011; Richter et al., 2017). Studies have shown that among all children who received early education interventions, children from high poverty families benefited the most (Berlinski et al., 2008) in school readiness and long-term academic success (Duncan & Magnuson, 2013; McCoy et al., 2017). Nevertheless, children in impoverished families were half as likely to attend ECE as their high-income counterparts (UNICEF , 2015)—a phenomenon known as "accessibility inequality." Accessibility inequality further exacerbates achievement gaps between social economic classes; for example, proxy measures of stunting and poverty indicate that 43% (250 million) of children younger than 5 in low SES families in developing countries are at risk of not reaching their developmental potential (Black et al., 2017). To address this accessibility challenge, many developing countries have devoted financial (Fiszbein & Schady, 2009; Handa & Davis, 2006), nutritional (Hoddinott et al., 2008), and educational (Baker-Henningham et al., 2009) resources to improve access to ECE for young children placed at high risk by poverty.

In the case of China, the ECE attendance rate increased seven-fold (from 11% to 77%) between 1976 and 2016 (Li, 2017). Nearly all children (98.4%) in urban areas in China have access to ECE (Ministry of Education China, Office of the Press Secretary, 2017), whereas only 30.1% of children from the poor and rural areas (mostly located in central and western China) have access to ECE (Zhang, 2015). The China Development Research Foundation (CDRF) of the State Council of China launched the One Village One Preschool (OVOP) initiative in 2009, with the aim of guaranteeing ECE access to all young children in rural China in next decade. The OVOP initiative will be introduced in detail below.

Value Added by One More Year in ECE

Building on success in improving ECE access for young children, scholars have proposed expanding exposure to ECE for children in poverty by providing ECE to younger children, in order to address the adversity due to poverty that begins early in life (Duncan et al., 2012). This proposal raises as a central question the possibility of dosage effects, specifically, the value added to developmental outcomes by 1 more year in ECE

(Reynolds et al., 2011). Many have found that greater participation in ECE contributed to better cognitive development for low SES children (Behrman et al., 2004; Dearing et al., 2009), but some studies found longer ECE duration to be associated with negative effects (Datta Gupta, & Simonson, 2010; Magnuson et al., 2007; Vandell et al., 2010). In a meta-analysis of dosage effects, Nores and Barnett (2010) concluded that programs shorter than 1-year yielded effect sizes of 0.2 SDs, whereas programs that lasted 3 years yielded only 0.3 SDs, rather than the 0.6 that might be expected. Indeed, many ECE programs produce nearly maximum effects with the first year of service (e.g. Gormley & Gayer, 2005). Although additional years of ECE can produce modestly higher effects than a single year (Belsky et al., 2007; Burchinal, 1999; Domitrovich et al., 2013; Skibbe et al., 2011; Wen et al., 2012; Zaslow, 2013), the size of the gains declines in subsequent years (Arteaga et al., 2014; Reynolds et al., 2011; Tarullo et al., 2013). This led scholars to argue that additional years of ECE may be redundant (Reynolds et al., 2011) as they do not provide unique increments to learning and development.

However, Jenkins et al. (2016), in an Oklahoma pre-K study, implicated the quality of the ECE experience by showing that children attending Head Start at age 3 and transitioning to academic pre-kindergarten programs at age 4 had better pre-reading skills than children who attended Head Start at both ages 3 and 4. From a policy perspective, prior dosage studies suggest investment in 1 year of ECE is the most cost-efficient approach unless multiple years of ECE were aligned to children's developmental progression while also adding new academic challenges in subsequent years.

The Long-Term Effect and the "Fadeout Phenomenon"

Numerous research studies have found ECE to have a long-term effect on participants' achievement in later years of schooling (Broekhuizen et al., 2016; Schweinhart, 2016) as well as their lifelong well-being (Barnett, 1998; Heckman et al., 2006). In studies of ECE's effect on primary school achievement, Berlinski et al. (2009) found a year of ECE was associated with an 8 percent increase in third grade test scores for a sample in Argentina. Hungi and Ngware (2018) found students from Uganda who attended ECE outperformed their counterparts who did not attend ECE in 6th grade math by 6 to 12 percent. In one study in Denmark, attending high quality preschool was reported to be associated with 0.01 to 0.02 SDs improvement in ninth grade test scores (Bauchmüller et al., 2014).

Growing evidence, however, has shown that most of the benefits from public ECE interventions remained short-term, fading by end of kindergarten or early grades of primary school (Puma et al., 2010; Weiland & Yoshikawa, 2013). Children who attended ECE tend to regress to the grand average by the third grade (Lipsey et al., 2013; Puma et al., 2012), the widely discussed "fadeout phenomenon" (Bailey et al., 2017; Bassok et al., 2018; Lipsey et al., 2015; Miller & Bassok, 2019). Although Jenkins et al. (2018) showed that professional supports to teachers substantially mitigated fadeout between kindergarten and first grade, more studies agreed that the benefits of preschool programs are sustained minimally or not at all. For example, in a study of the Tennessee's Voluntary Pre-K Program, Lipsey et al. (2015) showed that program participants

significantly outperformed nonparticipants at the end of the preschool year, but the performance of participants was matched or even exceeded by nonparticipants by the end of the second grade.

Prior studies have shown that the third grade in primary school is a critical year in a child's academic growth. It is the grade in which students from disadvantaged backgrounds, such as racial minority students (Fryer & Levitt, 2006) or students living in poverty (Hernandez, 2011), showed a significant achievement gap that is not explainable by observable covariates. It is also the grade in which teaching transitions, at least in Anglophone countries, from a focus on "learning to read" (simple decoding) to "reading to learn" (deep comprehension) (Fiester, 2010). Simple decoding skills are usually fully covered by the first 2 years of primary education, but deep comprehension skills are not readily trained in schools. They not only require a large amount of vocabulary that is not taught in school, but also an understanding of real-world concepts that rely on rich environmental and parental input (Snow et al., 1998), something that children in poverty often lack. Skill differences that were not usually tested in kindergarten or the primary grades may manifest themselves as an achievement gap in test scores by the third grade. For example, differences in the number sense skills in kindergarten or in the first grade may evolve to become the mathematical proficiency gap by the third grade (Jordan et al., 2010). Further, the achievement gap that emerges after the third grade has a lasting impact on students' high school performance and college enrollment levels (Lesnick et al., 2010).

No study, to the best of our knowledge, has examined the relationship between the duration of ECE enrollment and the magnitude of long-term fade out. One additional year of ECE may yield a small and marginal increment of knowledge or skills at the end of preschool but are these increments substantial enough to counteract the ECE fade out by the third or fourth grade? Or conversely, do the incremental knowledge or skills accumulated in an extra year of ECE fade out by the third or fourth grade? This is a question that interests both policy makers and educational theorists around the world. Unfortunately, there has not been any research on either the ECE duration effect or the fade out phenomenon in China; thus, one purpose of the current work is to explore whether duration effects and fadeout phenomena apply to the Chinese context.

One Village One Preschool: The Chinese Experience

In China, fast but concentrated economic growth has led to an expanded wealth gap between families from urban and rural areas. In 2016, the ratio of average per capita disposable income (PCDI) of urban to rural China was 2.72 (China State Statistics Bureau, 2017). As mentioned earlier, nearly all children (98.4%) in urban areas of China have access to ECE (Ministry of Education China Office of the Press Secretary, 2017), but only 30.1% of children from the poor and rural areas have access to ECE (Zhang, 2015).

Before the above mentioned OVOP initiative, there were only two types ECE centers that families in rural China could choose from: township public or township private ECE. Supported by public funds, public ECE centers in towns usually enjoy better physical environments, better teacher compensation, longer teacher retention, and higher

teacher-student ratios (1:18) than private centers (1:30). In contrast, private ECE centers in rural Chinese towns are inferior on several indicators of quality, including facilities and teacher preparation. Private centers mainly serve parents and children who are unable to access public centers due to household registration issues, financial and commuting concerns, or SES background. Almost all ECE centers were built in town centers, with very few in the neighboring villages that the towns administer. Regardless of the type of ECE, if a parent from a village wishes to send his/her child to ECE, he/she has to bring the child to the town center. Travel distance and the cost of tuition may deter families living in remote villages from participating.

Launched by China Development Research Foundation (CDRF) of the State Council of China in 2009, the One Village One Preschool (OVOP) initiative is the only nation-wide preschool program that aims to provide accessible free early education experience to children in remote villages. Our earlier study (Chen et al., 2019) has shown that OVOP has positive fadeout-resistant impacts on young children's academic achievement in elementary schools. In particular, whereas the test score rankings of students from the no-ECE and township private ECE groups decreased by the later years of elementary school, students from the OVOP group and the township public ECE group gained in rankings.

By 2020, OVOP will be incorporated into national policy that guarantees each poverty-stricken village with at least ten 3–6-year old children a minimum of one local, cost-free ECE center. To achieve full coverage of all poverty-stricken villages in China, about 100,000 OVOP centers will need to be built. As of 2018, 2,800 OVOP centers had been established in nine provinces. Cumulatively, 170,000 children have benefited from the OVOP program. It is estimated that enrolling each student costs 1700-2300 RMB (240–330 USD) annually

In the year 2012–2013, the OVOP initiative was in an early stage of development, and the total amount of funding was limited. Because the annual funding per classroom was the same across all centers, CDRF could not guarantee the 3-year program to all centers. Thus, 3-year centers were assigned to the poorest villages among the poverty-stricken villages, and 2-year centers were assigned to the rest. However, there were many village-level factors that determined the poverty metric of a village; if we compare the average family incomes between the children enrolled in the 2-year and 3-year centers, they were very similar. The 2-year and 3-year OVOP centers were nearly the same in teacher recruitment, school facilities, curriculum, and admissions policy, since they were given the same amount of annual funding per classroom.

The assignment of 2-year versus 3-year centers reflected the tradeoff between expanding the number of centers and prolonging the duration of attendance in the extreme poverty regions. As the scale of this national project continues to expand, this tradeoff remains an active topic of discussion among policy makers and scholars. It is thus of imminent importance to evaluate whether sustainable benefits of 1 more year of access to the OVOP program for children's long-term development offset the additional cost.

Research Questions

In this study, we obtained a complete sample of all fourth-grade students in Songtao, a national level poverty-stricken county in southwestern China. Each student had one of

five different ECE experiences: 3-year OVOP, 2-year OVOP, 3-year township public ECE, 3-year township private ECE, or no ECE. Using multinomial propensity score weighting to balance the covariates (e.g. gender, single child, poverty level, left-behind child, parent occupation and residential registration status) that may influence the ECE enrollment, we asked:

RQ1: Did the two OVOP groups achieve a higher score rankings than the no-ECE, township private or township public ECE groups by the end of the first grade in elementary school? In addition, did the 3-year OVOP group differ from the 2-year OVOP group?

RQ2: Assuming linear growth for each group from the first to the fourth grade, did the two OVOP groups achieve a steeper growth rate, compared to the other three groups from the first to the fourth grade in elementary school? In addition, did the 3-year OVOP group and the 2-year OVOP group have different growth rates?

RQ2: Assuming a discrete rate of growth for each group from the first to the fourth grade, did any of the five groups showed a decrease in rankings at the third grade? If there were declines in rank, did the 3-year OVOP group and the 2-year OVOP group have different magnitude of decline?

Method

Research Site

The participants of this study were from the Songtao Miao-Ethnic-Minority Autonomous County (Songtao) in Guizhou Province. Guizhou is one of the five poorest and most economically undeveloped provinces in China. It is located in the southwestern mountainous part of China. Guizhou is also one of the most ethnically diverse provinces in China; 37% of the population are members of ethnic minority groups.

Songtao County has a population of 740,000, and ethnic minority groups (primarily of the Miao ethnicity) account constitute more than 68.1% of the population.

Sample

The sample for this study is a complete dataset of all fourth-year elementary school students in Songtao county, Guizhou province, during the year 2015 to 2018 as documented by the county education department. All of the information used in the analysis in this article was provided by the country education department. The county education department in Songtao has information in each child's schooling record that includes schooling history (e.g., which type of kindergarten the child attended) and the final exam scores in each subject for each school-year starting from elementary school. The registration of this information is mandated by law in China. The analytic sample included all children for whom test scores at grade 4 were available. Due to the Compulsory Education Law in China, the attrition from grade 1 to grade 4 was less than 0.5%, suggesting that our sample represents something close to the entire Songtao 4th grade population in 2018.

The total sample was 23,775 students with the following preschool experiences: 1,712 of them did not attend any early childhood education center prior to elementary school (no-ECE group), 11,674 attended township public kindergartens (111 number of 3-year

Table 1. Descriptive statistics grouped by ECE types.

	3-year-OVOP	2-year-OVOP	Town Private	Town Public	No ECE
Sample size	1,656	3,813	4,921	11,674	1,712
Chinese-year-1					
mean	76.80	76.60	68.00	78.80	60.80
(sd)	(16.80)	(16.20)	(16.60)	(15.60)	(14.90)
Math-year-1					
mean	76.90	76.50	68.00	78.70	60.80
(sd)	(16.80)	(16.20)	(16.70)	(15.50)	(15.10)
English-year-1					
mean	76.80	76.50	68.00	78.70	60.80
(sd)	(16.80)	(16.20)	(16.70)	(15.50)	(15.10)
Sum-score-year-1					
mean	230.00	230.00	204.00	236.00	182.00
(sd)	(49.80)	(47.90)	(49.30)	(45.80)	(44.30)
Single child	44.60%	43.20%	47.80%	49.20%	42.00%
Left behind child	39.30%	38.60%	28.20%	17.10%	37.40%
Child w/ special need	6.76%	8.86%	4.57%	2.28%	8.05%
Poverty	12.70%	14.00%	2.23%	0.53%	14.10%
Mom w/ high school edu	23.30%	23.20%	39.70%	50.50%	24.80%
Mom township Hukou	11.40%	12.40%	26.30%	80.80%	14.30%
Dad township Hukou	7.25%	8.03%	22.70%	75.60%	6.52%
Han ethnicity	20.10%	18.10%	36.90%	39.30%	21.30%

Town-Public group), 4,921 attended private township kindergartens (109 number of 3-year Town-Private group), 3,813 attended OVOP for 2 years (200 number of 2-year-OVOP group, enrolled in 2013), and 1,656 attended OVOP for 3 years (100 number of 3-year-OVOP group, enrolled in 2012).

Table 1 shows the students' first year grades, as well as their background information. In general, students who went to township public kindergartens had higher first year grades in elementary school, and they were also more likely to come from higher SES backgrounds. Students who went to township private kindergartens had the second-best SES background on average, yet their first-year elementary school grades were the second-lowest among the five groups. Students who went to 3 or 3 years of OVOP had similar SES backgrounds to those who did not enroll in any ECE. These three groups were equivalently the lowest SES groups. The no-ECE group had the lowest first year grades in elementary school, whereas the 2-year and 3-year OVOP groups had equivalently the second-best first-year elementary school grades among the five groups.

The 3-year and 2-year OVOP groups are very similar on most variables except that the 3-year OVOP had more boys and fewer children with special needs than the 2-year OVOP group (marked † in Table 1). Compared to the resource-rich township public kindergartens, the 3-year OVOP group had fewer children from single child families, more left- behind children, a higher probability of poverty, lower maternal education, lower probability of township Hukou (residence permit), and less probability of Han ethnicity. Parents of public township kindergarten children had a significantly higher percentage of township Hukou than other groups.

OVOP: The Intervention

The OVOP intervention encompasses the physical buildings and facilities established as OVOP centers, the educational components (e.g. curriculum, classroom teaching, and

teacher training) of the OVOP program, and the policy and organizational operations (e.g. funding, policy making) of the OVOP initiative.

Buildings and Facilities

One critical challenge that children and their families living in rural villages faced was the commuting time for ECE. Few villages had ECE centers before the implementation of OVOP. Typically, parents and children needed to spend hours commuting to schools in townships, which is a significant investment of time and discourages participation. Therefore, OVOP aims to provide centers within reach of villages and neighboring families. The OVOP rents, repairs and decorates unused buildings that belong to elementary schools or village committees as venues. A CDRF survey indicates that 75% of OVOP centers are located within 30-minutes' walking distance to children's homes.

Besides the basic settings of classrooms, the survey found 65% of the centers have facilities to serve lunch and 32% have spaces where children can take afternoon naps. Some OVOP centers offer supplementary nutrition like afternoon snacks, and vitamin tablets.

Teacher and Professional Development

Local Education Bureaus manage the recruitment, selection, and orientation training of OVOP teachers. Typically, OVOP recruits graduates from local vocational schools or colleges who understand the dialects, ethnic languages, and culture of the area. CDRF's statistics from 2018 show that the average age of OVOP teachers is 28-years-old; 42.81% graduated from local teacher training vocational schools or colleges.

OVOP teachers have multiple professional development opportunities. Depending on the Education Bureau of each county, every school year the OVOP teachers may attend lectures, visit township public kindergartens, and study with experienced teachers from city kindergartens.

Class Organization and Curriculum

Most OVOP schools have a relatively small class size. The teacher-child ratio of OVOP schools is around 1:17, which is larger than the ratio of the low-quality township kindergartens and close to the ratio of well-resourced township public kindergartens. A significant feature of OVOP classes is that the children are of mixed-ages. More than half OVOP schools have only one class where children aged from 3 to 6-years-old play and learn together. Mixed-age ECE is not common in China, and there is no curriculum for mixed-age ECE available for use. OVOP teachers rely on the curriculum of the local township public kindergartens and teachers may not design their own lesson plans. No particular, well-designed curriculum has been developed for OVOP yet.

Sponsors and Funding of OVOP

Sponsors

The OVOP has three major sponsors: CDRF, county governments, and other charitable organizations. CDRF initiated the project, provided plans, raised seed funding, monitored the project, and sponsored an evaluation. Local governments signed contracts

with CDRF and collaborated with local villages and schools. They decided the number, location, and management of OVOP centers. The County Education Bureau is in charge of the daily schedule, teacher recruiting, professional training, and quality monitoring.

Funding

CDRF raised funds from enterprises, charitable organizations, and individuals to cover the basic cost of each OVOP. Typically, an OVOP with 20 students has operational costs of 30,000–50,000 RMB (about $4,500–$7,500) annually, which equals $225–$375 per student per year. CDRF pays for supplemental nutrition at a cost of 1.5 RMB per day ($0.22). CDRF also subsidizes each teacher 1,300 RMB ($194) and provides a 300 RMB ($45) commuting allowance per month. Every semester, a teacher has a training budget of 500 RMB ($75), while a classroom has an equivalent budget for toys and books. The monitoring and evaluation budget for each OVOP is 2,200 RMB ($328) per year.

Local governments provide each OVOP with 20,000 RMB ($3,000) to rebuild the campus, renovate, and decorate the classroom environment, and purchase furniture and playground equipment. Public funding from local governments provides extra subsidies to teachers and their professional development. Some counties provide more funding to cover CDRF's expenses. It is noteworthy that the competitive salary and professional benefits hardly attract township teachers to transfer to the villages because the living conditions in the village are much worse than in the township. CDRF adopted the "hire local and train local" strategy to recruit local villagers who recently graduated from high school. The salary and benefits were designed to attract local youth who would otherwise have left the village in search of better-paying jobs to stay and teach in the village.

The Development of OVOP

Phase 1: 2009–2011

The first two experimental field sites of the OVOP were Ledu, in Qinghai Province and Xundian, in Yunnan Province. CDRF started by renting unused village residences as venues, recruited qualified teachers, and provided at least two-half day ECE service in rural villages. During this period, an OVOP only opened for half days, to allow one teacher to rotate between two centers every workday.

Phase 2: 2012–2017

The rich experiences of OVOP in Ledu and Xundian have been extended to fifteen national level poverty-stricken counties in Guizhou (our study site—Songtao county—is in Guizhou province and joined the OVOP Initiative in Phase 2), Sichuan, Hunan, Xinjiang, Shanxi, and Gansu provinces.

Phase 3: 2017–Present Day

To respond to the national Targeted Poverty Alleviation Policy (Hu, 2016), the OVOP moved from being a social experiment to large-scale ECE policy for rural China. The

goal of OVOP during this phase was to improve the quality of OVOP, bring comprehensive ECE coverage to the most vulnerable 20% of children in China free of charge, and to build at least one OVOP center in every qualifying village.

Measures

Outcome

Academic achievement tests in Chinese, mathematics, and English were administrated at the end of the school year in the first to fourth grade. Within each grade, we calculated students' standardized score (z-score) for each subject as well as for the sum score. We term the standardized scores respectively, Chinese, Mathematics, English, and total. Each score was treated as the outcome variable in separate models. We only had access to students' final test scores, but not item scores. Thus, we could not calculate the internal reliability of the tests.

Predictors

In our sample, the major predictor was type of ECE education experience. Every elementary school documented students' early education experience during the admission process. There are five types: 3-year OVOP, which served about 7% of the Songtao ECE participants; 2-year OVOP, serving about 16%; township public kindergarten (public center), serving about 49%; township private kindergarten (private center), which served about 21%; or no early childhood education (no ECE) which was 7% of the sample.

OVOP

There were 5,469 students enrolled in OVOP Songtao before they entered elementary school. 1,656 of the students enrolled in OVOP from 2012 to 2015 (3-year OVOP group), and 3,813 of the students enrolled from 2013 to 2015 (2-year OVOP group).

Township Public ECE (Public Center)

In our sample, 11,674 students were enrolled in public centers. The township public ECE settings are funded and supervised by the local government. The Education Commission of Songtao consults on public centers' curriculum design and teaching quality. Typically, teachers working in public centers have more education than those working in private centers. They also enjoy higher salaries, better social benefits, and more professional development opportunities.

Township Private ECE (Private Center)

There were 4,921 students in the sample enrolled in private centers, which are run by private individuals or companies and have limited government support for financial needs or for the professional development of teachers. Teachers in private centers have lower salaries and fewer social benefits than do teachers at public centers.

No ECE
In the sample, 1,712 students had no early childhood education experience.

Control Variables
The National Compulsory Education Information Gathering System (NCEIGS) collects students' basic demographic and family information (gender, *hukou*). CDRF sent out surveys to collect additional background information, including left-behind status (if both parents were working in urban regions and not living with the child), special needs, poverty, mother's education. Based on child development theories, we selected critical control variables that matter to young children's academic achievement and that may influence the choice of different types of ECE settings. Information on these children's demographic features, family, and parental characteristics was collected during September 2018.

Girl
The dummy variable of gender. We code girl = 1.

Only Child
This dummy variable = 1 means the student is the only child in the family.

Left-Behind Child
The left-behind child = 1 means the student remains in a rural region while both parents have left the region for work in urban areas. Left-behind children are taken care of by relatives, usually by grandparents or family friends, who stay in the rural areas.

Special Needs
In our sample there are three types of special-needs children: (a) those with physical disabilities, (b) intelligence challenges, and (c) orphans. The special needs = 1 means the child was classified as belonging to at least one of the three categories.

Poverty
Poverty is a dummy variable that indicates the income of the family. Poverty = 1 means the family total annual income is lower than the Chinese national poverty line (lower than 320 USD); Poverty = 0 means the family's total annual income is above the poverty line; it is noteworthy that the average annual individual income was lower than 320 USD in Songtao county as of 2012.

Guardian as Parents
This dummy variable was scored as 1 if the child was living with at least one of his or her parents.

Township Hukou: Mother

A score of 0 on this dummy variable indicated that the mother of the child had an agriculture household registration (agriculture *hukou*) and was working in the field of farming. A score of 1 on this variable indicated that the mother has township Hukou. The Hukou system in China is a family registration program that can be considered a domestic passport. People with agricultural Hukous are not given access to social welfare in townships or cities (i.e., public services, healthcare, jobs). Each year there is a strictly enforced quota on the number of hukous that may be converted. Only people who have agricultural hukous are allowed to own land for farming (Zhou, 2020). Predictably, the township public and township private groups were more likely to have township hukous than the OVOP group.

Township Hukou: Father

Defined and coded in the same manner as the Hukou for mother.

Mother's Education

The education degree of the mother.

Analytic Approach

A critical challenge of this study was selection bias. Parents who chose different ECE types differ in ways that may also influence child development. A direct application of OLS regression cannot parse out the confounding effects. More problematically, OLS could be heavily influenced by extreme values, especially when sample sizes differ greatly and covariates overlap poorly between groups (see Ho et al., 2007). Thus, we used multinomial propensity score weighting (MPSW) techniques to balance the pretreatment variables between groups and to minimize bias. MPSW estimation approaches reconstructed the observational data to mimic experimental data and equated multiple groups based on the probability of receiving the treatment given pretreatment variables. By balancing pretreatment variables, MPSW techniques can reduce selection bias in nonexperimental studies.

Using multinomial propensity score weighing, we have reached a good balance in each of the covariates between every pair of groups (p value > 0.05 in all cases). The balance check shows that before weighting, there were significant differences on many covariates between groups; and after weighting all mean difference p-values or KS test p-values were bigger than 0.05.

We compared the model before and after applying multinomial propensity score weighting, the results are identical in terms of the ranking and growth of each group. In this study, we only reported the result after weighting.

The students in our sample are clustered in schools. Please note that we did not use multilevel regression although it is typically the preferred choice for a hierarchical data structure. In this case, elementary school information was a post-treatment variable (meaning that children went to elementary school after receiving the ECE treatment). Variables that are in part a consequence of the treatment variable should never be

controlled for when estimating a causal effect (Ho et al., 2007; Rosenbaum, 1984). Controlling for the consequences of a causal variable can severely bias a causal inference. This "post-treatment bias" is common in social science studies (King and Zeng, 2007). It is not advisable to control post-treatment variables after weighting or matching, either as controlled variables or as multilevel clustering (e.g., Chen et al., 2019; Montgomery et al., 2018), because it would re-introduce post-treatment bias. In our estimation, that would undermine the effect of using the IPW bias reduction (see detailed statistical explanation in Rosenbaum, 1984). To deliver conservative estimates in nested data structures without using multilevel modeling, we reported robust standard errors to protect against Type I error inflation.

Results

Because we were not only interested in the students' first year elementary school grades (the intercept) but also the slopes of their longitudinal trajectory, we adopted a longitudinal growth model (Singer & Willett, 2003) in our estimation of the ECE effects for the Z-score for each subject (and also for the total). We first assumed that the effect of time (years in our case) was linear. In the alternative models, we relaxed the linear assumption and treated years as discrete dummy variables. Both specifications are presented in Table 2, where L indicates linear year, and D indicates discrete year. The reference ECE type is No-ECE. When year was treated as discrete, the reference year is the first year. We also allowed the effect of year to vary by ECE types; therefore, we included interaction effects between year and ECE types in our model.

First, by examining the intercept and the main effects of ECEs, we found the Town-Public group had the highest first year grades in all subjects, followed by the 3-year-OVOP, 2-year-OVOP, Town-Private, and lastly No-ECE groups. The first-year grades of the 3-year-OVOP and the 2-year-OVOP groups were not statistically significantly different from one another. Town-Private had significantly lower first year grades than Town-Public and the two OVOP groups but significantly higher than the No-ECE group. This ranking can be seen from the intercepts in Figures 1 or 2.

Second, the growth slopes were complicated by the significant effect of year and the significant interaction effects between year and ECE type. It was best summarized by Figure 1 for linear growth and Figure 2 for discrete growth. Because all the subjects had nearly identical intercepts and growth trends, we only illustrated the total scores as the outcomes in Figures 1 and 2.

Focusing on the linear specification of year, in Figure 1, we see nearly flat growth in the Town-Public group. Town-Public remained the top group across the first 4 years of elementary school. The 3-year-OVOP group had a positive slope. Its average sum score increased by 0.02 standard deviations each year. Additionally, by the 4th grade, they had decreased their gap with the Town-Public group from 0.14 SD to 0.05 SD.

By contrast, the 2-year-OVOP group had a negative slope of 0.03 SD and decrease in Z-score per year. Although the 3-year-OVOP and 2-year-OVOP had nearly identical grades in the first year (0.02 SD), the 2-year-OVOP group trailed behind the 3-year-OVOP group by as much as 0.18 SD by the fourth year. Town-Private had positive slope similar to the magnitude of the 3-year-OVOP group; its total Z-score grew by

Table 2. Predicting elementary grades using multinomial propensity score weighted growth models.

Fixed effects:	M(Total_L) Estimate	se	M(Total_D) Estimate	se	M(Chinese_L) Estimate	se	M(Chinese_D) Estimate	se	M(Math_L) Estimate	se	M(Math_D) Estimate	se	M(English_L) Estimate	se	M(English_D) Estimate	se
Intercept	−0.491	0.017***	−0.507	0.017***	−0.481	0.017***	−0.501	0.016***	−0.492	0.017***	−0.505	0.016***	−0.481	0.017***	−0.495	0.016***
year	−0.018	0.001***			−0.019	0.002***			−0.016	0.002***			−0.019	0.002***		
year2			−0.043	0.004***			−0.033	0.005***			−0.043	0.005***			−0.051	0.005***
year3			−0.004	0.004			−0.003	0.005			−0.003	0.005			−0.005	0.005
year4			−0.074	0.004***			−0.073	0.005***			−0.068	0.005***			−0.078	0.005***
TownPublic	1.049	0.031***	1.047	0.031***	1.042	0.031***	1.038	0.030***	1.032	0.031***	1.035	0.030***	1.030	0.031***	1.025	0.030***
3-Year-OVOP	0.910	0.026***	0.926	0.026***	0.911	0.026***	0.927	0.026***	0.904	0.026***	0.910	0.026***	0.880	0.026***	0.903	0.026***
2-Year-OVOP	0.931	0.020***	0.921	0.020***	0.918	0.020***	0.915	0.020***	0.922	0.020***	0.905	0.020***	0.916	0.020***	0.905	0.020***
TownPrivate	0.370	0.017***	0.414	0.017***	0.368	0.017***	0.416	0.017***	0.366	0.017***	0.407	0.017***	0.362	0.017***	0.402	0.017***
gender	0.011	0.012	0.011	0.012	0.008	0.012	0.008	0.012	0.013	0.012	0.013	0.012	0.012	0.012	0.012	0.012
Han ethnicity	−0.014	0.013	−0.014	0.013	−0.014	0.013	−0.014	0.013	−0.013	0.013	−0.013	0.013	−0.013	0.013	−0.013	0.013
momedu	−0.001	0.013	−0.001	0.013	−0.001	0.013	−0.001	0.013	−0.001	0.013	−0.001	0.013	−0.001	0.013	−0.001	0.013
dadhukou	−0.007	0.017	−0.007	0.017	−0.007	0.016	−0.007	0.016	−0.006	0.016	−0.006	0.016	−0.007	0.016	−0.007	0.016
momhukou	0.005	0.015	0.005	0.015	0.002	0.015	0.002	0.015	0.008	0.015	0.008	0.015	0.004	0.015	0.004	0.015
poverty	−0.009	0.024	−0.009	0.024	−0.009	0.023	−0.009	0.023	−0.008	0.023	−0.008	0.023	−0.009	0.023	−0.009	0.023
singlechild	0.001	0.012	0.001	0.012	0.001	0.012	0.001	0.012	0.002	0.012	0.002	0.012	−0.002	0.012	−0.002	0.012
specialneed	0.038	0.025	0.038	0.025	0.040	0.025	0.040	0.025	0.037	0.025	0.037	0.025	0.038	0.025	0.038	0.025
year × TownPublic	0.015	0.002***			0.014	0.002***			0.015	0.002***			0.015	0.002***		
year × 3-Year-OVOP	0.038	0.002***			0.032	0.002***			0.038	0.002***			0.043	0.002***		
year × 2-Year-OVOP	−0.012	0.002***			−0.012	0.002***			−0.012	0.002***			−0.012	0.002***		
year × TownPrivate	0.043	0.002***			0.041	0.002***			0.042	0.002***			0.044	0.002***		
year2 × TownPublic			0.084	0.006***			0.086	0.008***			0.066	0.007***			0.098	0.007***
year3 × TownPublic			−0.010	0.006~			−0.015	0.008*			−0.001	0.007			−0.013	0.007~
year4 × TownPublic			0.081	0.006***			0.079	0.008***			0.072	0.007***			0.088	0.007***
year2 × 3-Year-OVOP			0.102	0.005***			0.079	0.007***			0.119	0.007***			0.106	0.007***
year3 × 3-Year-OVOP			0.082	0.005***			0.061	0.007***			0.099	0.007***			0.083	0.007***
year4 × 3-Year-OVOP			0.133	0.005***			0.112	0.007***			0.132	0.007***			0.150	0.007***
year2 × 2-Year-OVOP			0.057	0.005***			0.041	0.007***			0.069	0.007***			0.059	0.007***
year3 × 2-Year-OVOP			−0.172	0.005***			−0.181	0.007***			−0.155	0.007***			−0.173	0.007***
year4 × 2-Year-OVOP			0.036	0.005***			0.035	0.007***			0.034	0.007***			0.039	0.007***
year2:TownPrivate			0.040	0.005***			0.020	0.007**			0.046	0.007***			0.052	0.007***
year3:TownPrivate			0.089	0.005***			0.086	0.007***			0.085	0.007***			0.091	0.007***
year4:TownPrivate			0.127	0.005***			0.116	0.007***			0.128	0.007***			0.132	0.007***
Random effects:	Variance		Variance		Variance		Variance		Variance		Variance		Variance		Variance	
Intercept	0.825		0.826		0.797		0.798		0.797		0.7982		0.797		0.7974	
Residual	0.148		0.141		0.2550		0.2473		0.250		0.2435		0.251		0.2438	

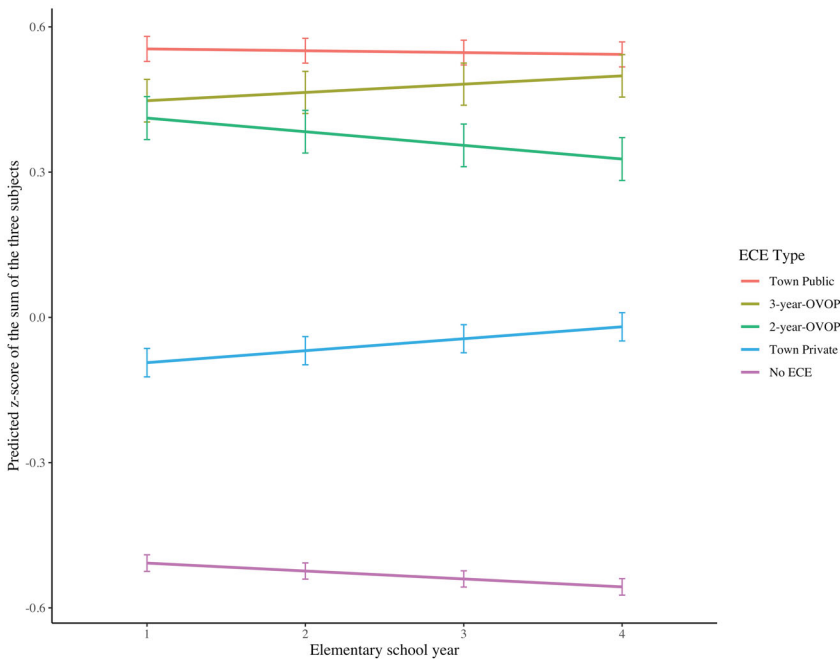Notes: *** p < 0.001, ** p < 0.01, * p < 0.05, ⊠ p < 0.08 after FDR adjustment

Figure 1. Predicted Z-score, with 95% confidence intervals, of the sum of three subjects (Chinese, math and English) as a function of linear elementary school year and ECE type.

0.02 per year. Although they were still behind the 2-year-OVOP group by 0.28 SD, their group average sum score was only 0.01 SD from the grand mean. The No-ECE group had a negative slope, decreasing by 0.02 SD per year.

We will shift our focus to the discrete specification of year as shown by Figure 2, where we see that most "eventful" year was grade 3. According to the discrete time model, the 3-year-OVOP group had surpassed the Town-Public group at year three, although they fell behind again in year four. The 2-year-OVOP group's Z-score dropped by 0.23 SD in comparison to its second year. It increased this gap from 3-year-OVOP, from only 0.05 SD by the second year to 0.26 SD. The Town-Private group increased by 0.05 SD from the second to the third year. The No-ECE group's average third year Z-score was the same as their first year; however, their second- and fourth-year measurements had significantly lower Z-scores by 0.04 and 0.07 SD. The group means for Figure 2 can be found in Table 3 in the online supplementary materials.

In summary, Town-Public consistently achieved the highest grades on average, and its Z-score remained flat. The 3-year-OVOP and 2-year-OVOP groups started off at the same grade in the first year, but the 3-year-OVOP group grew steadily while the 2-year-OVOP group declined. The decline of the 2-year-OVOP was the steepest in grade 3. Town-Private grew at a similar pace to that of the 3-year-OVOP group, with its largest gain occurring in the third year. The No-ECE group ranked last in the first year, and its ranking kept declining. The relative intercept and trend among the five groups remained nearly identical for each of the individual subjects.

Although the 3-year-OVOP and 2-year-OVOP groups shared a similarly low SES background with the No-ECE group, they out-scored the No-ECE group by 0.95 and
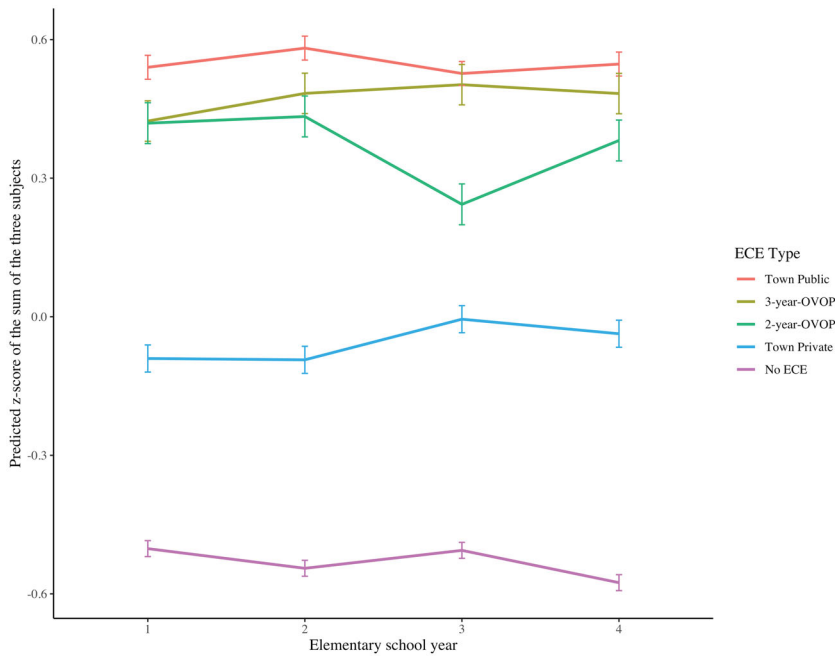
Figure 2. Predicted Z-score, with 95% confidence intervals, of the sum of three subjects (Chinese, math and English) as a function of discontinuous elementary school year and ECE type.

0.92 SD respectively by the first year of elementary school. They out-scored the No-ECE by 1.06 and 0.88 SD respectively by the fourth year. Although Town-Private had a similarly high SES background to that of the Town-Public group, and was much higher than the two OVOP groups, the grades of Town-Private children consistently fell behind the two OVOP groups. This pattern was confirmed in analyses conducted both with and without propensity score weighting.

Discussion

Motivated by the new direction of early childhood education policy in China, which seeks a balance between expanding scale and lengthening service duration, this study aimed to evaluate the additive effect of 1 more year of the OVOP program on children's intercept, growth rate, and potential fade out rate in elementary school. In the context of China, this is the first study that compared the 3-year OVOP and 2-year OVOP programs; in the worldwide context, this is the first study to examine the relationship between ECE duration and long-term fade out. To examine these research questions, we collected a complete sample from fourth graders from Songtao County in 2018. We used individual growth modeling with multinomial propensity score weighting approach to compare students' Chinese, mathematics, and English test score rankings and growth among five ECE conditions: 3-year OVOP, 2-year OVOP, township public, township private or no-ECE.

Our findings showed that (a) 3-year and 2-year OVOP groups achieved similar rankings in all subjects by the end of first grade, which was lower than township public ECE

group but higher than the township private or no-ECE group. (b) the 3-year OVOP group had a positive growth rate, but the 2-year OVOP group had a negative growth rate. By the third or fourth grade, the 3-year OVOP group had matched the township public ECE group, and the 2-year OVOP group, although still ranked above the township private or no-ECE groups, lagged behind the 3-year OVOP group by about 0.2 SDs. (c) when individual growth was modeled to be discrete (rather than linear), we discovered that the 2-year OVOP group had the largest magnitude of fade out at the end of the third grade, whereas the 3-year OVOP group maintained, and even slightly improved, its ranking. We discuss three major findings below.

ECE Effect

If we only focus on the first-grade results, our findings concurred with the prior literature in two aspects: a positive ECE effect and a negligible dosage effect. Focusing on the ECE effect, we found that either type of OVOP experience would place the students, on average, 0.3 SDs above the grand mean. This is not only 0.7 SDs above the no-ECE group, but also 0.4 SDs above the township private ECE group. Considering the lack of resources, experience, and high poverty environment in the OVOP program as compared to township public ECE centers, it was not surprising that students in the OVOP groups did not achieve test scores as high as the township public ECE group by the end of the first grade. It was, however, somewhat surprising that OVOP groups consistently outperformed the township private ECE group in all four grades, considering that the township ECE groups have been operating for many more years than OVOP centers and that the town centers are relatively more affluent and more stimulating than the village locations. Three factors may explain the advantage of OVOP centers over township private centers: (a) OVOP centers were located in the same villages where the children live. Several children who went to the township centers had transferred to the OVOP centers when they found the OVOP was a more convenient and safe choice with the walking-distance commute, as compared to the daily long-distance travel between villages and town centers. Vernon-Feagans et al. (2012) showed that exhausting daily commuting routines may be associated with negative developmental effects. (b) OVOP centers offer teachers more competitive salaries than the township private ECE centers, and OVOP centers also offer government staffing status and government subsidies that township private centers do not offer. Thus, OVOP centers may retain teachers for a longer period of time, whereas township private centers are disrupted by high teacher turn-over. (c) OVOP centers have a higher teacher-student ratio than the township private centers, so OVOP children received more attention from teachers in a less crowded or chaotic classroom environment.

Dosage Effect

Regarding the dosage effect as observed by the end of first grade, we found that the 3-year OVOP group did not produce significantly more benefit than the 2-year OVOP group. This finding agreed with most recent research in the dosage effect. Specifically, that additional years in OVOP did not increase the benefit incrementally (e.g. Arteaga et al., 2014; Ramey

et al., 2000; Tarullo et al., 2013). This is mainly because, as argued by Jenkins et al. (2016), most of the public ECE interventions were not designed to build on (academic) gains and spending 1 more year would not introduce incrementally new knowledge or skills to the children. Reynold et al. (2011) once suggested that one additional year of ECE intervention was, in most cases redundant, a pessimistic appraisal of the dosage effect. Had we only observed the first-grade outcomes, we would have likely accepted this appraisal.

Nevertheless, when we expanded our view beyond the first grade to include the first to fourth grades trajectories, based on either the linear or discrete model, we came to a very different conclusion and a much more optimistic evaluation of the dosage effect. First of all, the linear model showed that the 3-year OVOP group had a positive slope of 0.02 SDs per year, but the 2-year OVOP group had a negative slope of 0.03 SDs per year. By the end of the fourth year, the 3-year OVOP had nearly caught up with the township public ECE group, whereas the 2-year OVOP group trailed behind the 2-year OVOP group by 0.18 SDs. It was surprising that the two OVOP groups that started off the first grade with identical rankings in all subjects would diverge from one another by 0.18 SDs in 4 years, with a separation rate of 0.045 SDs per year.

To attain a better understanding of the different trajectories between 3-year and 2-year OVOP groups, we will now shift our attention to the discrete model. The discrete model presented one salient message: the 2-year OVOP group had a large magnitude (0.2 SDs) of fade out at the end of the third grade, whereas the 3-year OVOP group maintained its ranking and improved slightly. The fade out trend observed in the 2-year OVOP group confirmed the results of prior studies that show the benefit of public ECE intervention tends to diminish over the long run, and this decline is the most noticeable in the third grade.

As explained in the introduction, the fade out phenomenon may be attributable to the heavy academic focus starting from grade three. A great amount of academic learning relies on out-of-class knowledge or fundamental cognitive skills that are not usually measured in preschools or early years in elementary schools, which could explain the contrasting fade out patterns of the 3-year and 2-year OVOP groups. One additional year in OVOP, although it did not prepare students with incremental knowledge to help them excel in tasks encountered in the first grade, may prepare the students with basic, yet lasting skills that are developmentally appropriate for 4-year old children, such as (a) world knowledge or vocabulary, (b) phonological awareness, (c) number sense skills, or (d) self-regulation skills. If children enroll in OVOP at the age of 5 (2-year OVOP group), they likely have spent their time as 4-year-olds without any formal educational input (like no-ECE children) since 38.6% of them were left-behind children cared for by guardians with very limited education. None of these abovementioned skills were measured in the first or second grade countywide tests; they were nevertheless fundamental to the advanced academic learning that started beginning in the third grade: world knowledge and vocabulary are key predictors of reading comprehension (Snow, 2006); number sense skill is a key predictor of mathematical operation (Jordan et al., 2010); phonological awareness is a key predictor of English decoding skills (e.g. Torgesen et al., 1997); and self-regulation is fundamental to long-term learning habits and resilience (Nota et al., 2004). For example, students can be slow but still accurate in identifying the Pinyin (phonetic aids for Chinese character reading) in the test in the first or second

grade. As students reach the third grade, they start to "read to learn" which requires proficient and fast-processing phonological skills to read without the aid of Pinyin; students who had more exposure to early literacy skill training in kindergartens may excel in this stage. In other words, it is possible that 1 additional year of the OVOP program yielded "dormant" benefits that were not measured in the first 2 years of elementary school where the focus of learning or testing was not on deep comprehension; however, these dormant benefits could develop into critical and observable advantages in various subjects as the focus of learning or testing shifted to deep comprehension.

Admittedly, we did not measure the intermediate skills that we speculate are at work, and our speculation remains to be examined in future studies. What we can conclude is that if we observe the course of a 4-year trajectory, 1 additional year of OVOP made a huge difference in student learning: it shifted OVOP students from a course of gradual fade out to sustainable growth that matched that of the children who received the highest quality ECE programs (the township public ECE group). The magic ingredient of the one additional year of ECE remains to be studied in the future.

Future studies should conduct more in-depth explorations of the issue of "dosage." In the current study, we used "year" as a general time unit to compare the dosage effects between 2-year OVOP and 3-year OVOP. Future studies could go deeper, adding the dimension of early childhood education quality (e.g., dosage of curriculum implemented, amount of teacher-child interaction, and children's learning opportunities) to the calculation of dosage. This information will help us understand why and how the extra year of preschool makes a difference in future academic outcomes.

Limitations

There were numerous limitations to this study.

   a. As mentioned above, we did not measure or track students' basic skills other than subject test scores in preschool or in elementary school. Therefore, we could not examine the dormant skills that contributed to the long-term bifurcation between the 3-year and 2-year OVOP groups.
   b. we did not observe the classroom practices in the OVOP programs. Teachers have nearly absolute freedom to dictate the content, the pedagogy and the activities, so we could not tease out the unique events that occurred during the additional year of OVOP. The research team is undergoing curriculum design to improve the curriculum quality and consistency. It is also experimenting with different curriculum and pedagogical features to accommodate the special needs of children from poverty-stricken villages.
   c. Songtao is only one country in southeastern China. Although we collected a complete sample of the county, we could not generalize our findings to all poverty-stricken counties in China. Nevertheless, there are reasons to believe that the Songtao case may represent a typical comparison between OVOP and other types of ECE centers:

Our study site—Songtao County—is one of the counties in Guizhou province officially categorized as poverty-stricken. Guizhou province was one of the first provinces

to participate in the OVOP initiative (Phase 2) (the other province was Qinghai). Guizhou province has by far the most comprehensive coverage of OVOP centers in poverty areas. CDRF intended the Guizhou centers to serve as examples that other provinces could learn from and emulate. Therefore, our study sites were designed to represent the standard for OVOP centers in China.

The government of each province in China gave the same amount of budget per pupil to each public kindergarten located in the province. According to Statistical Announcement on the Implementation of National Education Funds in 2018 (Ministry of Education of China, 2019), per pupil budget of kindergarten student of Guizhou Province was identical to the national average (7992 RMB and 7671 RMB). Other provinces ranked in the bottom 30% of GDP in China had similar per kindergarten student expenditure with Guizhou. Thus, all public kindergartens in each poverty-stricken county should have similar resources. In addition, the curriculum guidelines for all public kindergartens are the same across China. Each province can select one curriculum (including textbooks) from a list approved by the national Ministry of Education. The public kindergartens in Songtao County use the similar curricula as other poverty-stricken counties in China.

Private kindergartens in China have three curricular options: (a) choose from the list of curricula that are approved by the national Ministry of Education, (b) do not use any curriculum (very rare), and (c) design their own curriculum. In poverty-stricken areas in China, most teachers in private kindergartens are not equipped to design curriculum by themselves. Most of these kindergartens purchase the cheapest curricula from the approved list. In addition, the amount of tuition that a private kindergarten can charge is set by Administration of Commodity Prices based on the local family income. Thus, the private kindergartens in the poverty areas all have about the similar amount of per pupil budget.

This study replicated our earlier finding of the effectiveness of OVOP program (the earlier study only examined the OVOP effect and did not examine dosage effect) in a separate OVOP site in northwestern China (Chen et al., 2019). Thus, it incrementally strengthened our confidence in the external validity of the OVOP effectiveness. As the OVOP initiative is expanding to more counties, we will be able to evaluate the generalizability of the OVOP effect, the dosage effect, and their relationships with the fade out phenomenon.

We did not have information regarding elementary school characteristics in this study. Not controlling for the elementary school information should not be considered a major threat to the validity of our analysis because, as explained earlier, elementary school enrollment is a post-treatment event in our study context, and there is established literature in statistics that shows controlling for post-treatment information may increase bias in the estimation of treatment effect. Nevertheless, collecting elementary school information would allow us to examine if, and to what extent, students' growth can be attributed to elementary school heterogeneity. While acknowledging this limitation, we argue that the elementary schools in a poverty-stricken county can be considered homogeneous.

In almost all poverty-stricken counties in China, there are not any elementary schools at the village level. Elementary schools are only allocated to the township

centers (this was also the problem with preschool education before the OVOP movement—that all kindergartens were allocated to the townships, not any to villages). Songtao County has 28 townships, each consisting of about 20 villages. Each township has 2 or 3 elementary schools that admit all students from the township center and the villages affiliated with the township non-selectively. In other words, students from villages go to the elementary schools in the nearest township centers. All townships that equally share (weighted by population) the county-wide fiscal revenue and subsidiaries from the central government have roughly the same level of poverty. The teachers in elementary schools are trained and certified by exactly the same teacher training program organized by the county level education department, in the county capital. These teachers are randomly assigned to elementary schools in each township within the county. Teachers rotate between elementary schools every 3-5 years. Therefore, all elementary schools (and middle schools) are designed to have equal amounts of resources and to provide an equivalent level of quality in education. Thus, we suggest that elementary schools in our study should be considered to share homogeneous characteristics. It is noteworthy that this egalitarian policy is only applied to poverty-stricken counties in China; non-poverty counties have a wide divide in school quality associated with the SES distribution.

This study did not explore the impact of elementary schools' quality on ECE effects' longevity for two reasons: limited access to relevant data and to avoid post-treatment bias. In the future, researchers should consider addressing whether fadeout, maintenance or recovery of kindergarten impacts is related to elementary school quality by collecting data on the quality of the elementary schools attended by children in the different ECE groups.

Our multinomial propensity score weighting approach was intended to balance the factors that influenced the treatment assignment mechanism; however, it did not address all possible factors, thus we could not exclude the possibility self-selection bias and could not make a strong causal argument. We will conduct randomized control trials to make a robust evaluation of the causal impact of the OVOP program.

Conclusion

The OVOP initiative will be written into China's the national policy by the next 5 years. Building upon the initial success of the 2,800 OVOP centers, nearly 100,000 new OVOP centers will be built in the next decade. In addition to expanding the quantity of centers, scholars, practitioners, and policy makers are also interested in enhancing the quality and long-term impact of OVOP programs to help children develop to their full potential. One possible approach to enhance the quality of the program is to extend the duration of the program for each child from 2 to 3 years, which will also increase the cost of supporting each OVOP center. Thus, careful evaluation of the long-term benefits of 1 additional year of the OVOP program will provide important empirical evidence as to the cost-efficiency of the program. Our study strongly suggested that the evaluation of the cost-efficiency of public ECE programs in China should look beyond the early years of elementary school, take into consideration the long-term trajectory, and consider sustainable benefits relative to the fade out trend as one of the key indicators of

effectiveness. From this perspective, our study has shown that 1 additional year of OVOP program indeed brought children sustainable long-term benefits.

Funding

ORCID

Si Chen ⓘ http://orcid.org/0000-0003-2025-1547

References

Arteaga, I., Humpage, S., Reynolds, A. J., & Temple, J. A. (2014). One year of preschool or two - is it important for adult outcomes? Results from the Chicago longitudinal study of the child-parent centers. *Economics of Education Review*, *40*, 221–237. https://doi.org/10.1016/j.econedurev.2013.07.009

Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*, *10*(1), 7–39.

Baker-Henningham, H., Walker, S., Powell, C., & Gardner, J. M. (2009). A pilot study of the Incredible Years Teacher Training programme and a curriculum unit on social and emotional skills in community pre-schools in Jamaica. *Child: care, Health and Development*, *35*(5), 624–631.

Barnett, W. S. (1998). Long-term cognitive and academic effects of early childhood education on children in poverty. *Preventive Medicine*, *27*(2), 204–207.

Bassok, D., Gibbs, C. R., & Latham, S. (2018). Preschool and children's outcomes in elementary school: Have patterns changed nationwide between 1998 and 2010? *Child Development*, *90*, 1875–1897.

Bauchmüller, R., Gørtz, M., & Rasmussen, A. W. (2014). Long-run benefits from universal high-quality preschooling. *Early Childhood Research Quarterly*, *29*(4), 457–470. https://doi.org/10.1016/j.ecresq.2014.05.009

Behrman, J. R., Cheng, Y., & Todd, P. E. (2004). Evaluating preschool programs when length of exposure to the program varies: A nonparametric approach. *Review of Economics and Statistics*, *86*(1), 108–132. https://doi.org/10.1162/003465304323023714

Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., Owen, M. T, & NICHD Early Child Care Research Network. (2007). Are there long-term effects of early child care? *Child Development*, *78*(2), 681–701. https://doi.org/10.1111/j.1467-8624.2007.01021.x

Berlinski, S., Galiani, S., & Gertler, P. (2009). The effect of pre-primary education on primary school performance. *Journal of Public Economics*, *93*(1–2), 219–234. https://doi.org/10.1016/j.jpubeco.2008.09.002

Berlinski, S., Galiani, S., & Manacorda, M. (2008). Giving children a better start: Preschool attendance and school-age profiles. *Journal of Public Economics*, *92*(5–6), 1416–1440. https://doi.org/10.1016/j.jpubeco.2007.10.007

Black, M. M., Walker, S. P., Fernald, L. C. H., Andersen, C. T., DiGirolamo, A. M., Lu, C., McCoy, D. C., Fink, G., Shawar, Y. R., Shiffman, J., Devercelli, A. E., Wodon, Q. T., Vargas-Barón, E., & Grantham-McGregor, S. (2017). Early childhood development coming of age:

science through the life course. *The Lancet*, *389*(10064), 77–90. https://doi.org/10.1016/S0140-6736(16)31389-7

Broekhuizen, M. L., Mokrova, I. L., Burchinal, M. R., Garrett-Peters, P. T., & Family Life Project Key Investigators. (2016). Classroom quality at pre-kindergarten and kindergarten and children's social skills and behavior problems. *Early Childhood Research Quarterly*, *36*, 212–222.

Burchinal, M. R. (1999). Child care experiences and developmental outcomes. *The ANNALS of the American Academy of Political and Social Science*, *563*(1), 73–97. https://doi.org/10.1177/000271629956300105

Chen, C., Haduong, P., Brennan, K., Sonnert, G., & Sadler, P. (2019). The effects of first programming language on college students' computing attitude and achievement: A comparison of graphical and textual languages. *Computer Science Education*, *29*(1), 23–48. https://doi.org/10.1080/08993408.2018.1547564

Chen, S., Zhao, C., Cao, Y., Chen, C., Snow, C. E., & Lu, M. (2019). Long-term effects of China's One Village One Preschool Program on elementary school academic achievement. *Early Childhood Research Quarterly*, *49*, 218–228. https://doi.org/10.1016/j.ecresq.2019.06.010

China State Statistics Bureau. (2017, July 07). *Data shows that the income gap among residents in China is reduced.* http://money.people.com.cn/n1/2017/0707/c42877-29388944.html

Datta Gupta, N., & Simonsen, M. (2010). Non- cognitive child outcomes and universal high quality child care. *Journal of Public Economics*, *94*(1–2), 30–43. https://doi.org/10.1016/j.jpubeco.2009.10.001

Dearing, E., McCartney, K., & Taylor, B. A. (2009). Does higher quality early child care promote low- income children's math and reading achievement in middle childhood? *Child Development*, *80*(5), 1329–1349.

Domitrovich, C. E., Morgan, N. R., Moore, J. E., Cooper, B. R., Shah, H. K., Jacobson, L., & Greenberg, M. T. (2013). One versus two years: Does length of exposure to an enhanced preschool program impact the academic functioning of disadvantaged children in kindergarten? *Early Childhood Research Quarterly*, *28*(4), 704–713. https://doi.org/10.1016/j.ecresq.2013.04.004

Duncan, G. J., & Magnuson, K. (2013). Investing in preschool programs. *The Journal of Economic Perspectives: a Journal of the American Economic Association*, *27*(2), 109–132.

Duncan, G. J., Magnuson, K., Kalil, A., & Ziol-Guest, K. (2012). The importance of early childhood poverty. *Social Indicators Research*, *108*(1), 87–98. https://doi.org/10.1007/s11205-011-9867-9

Engle, P. L., Fernald, L. C., Alderman, H., Behrman, J., O'Gara, C., Yousafzai, A., de Mello, M. C., Hidrobo, M., Ulkuer, N., Ertem, I., & Iltus, S. (2011). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *The Lancet*, *378*(9799), 1339–1353. https://doi.org/10.1016/S0140-6736(11)60889-1

Fiester, L. (2010). *Early warning! Why reading by the end of third grade matters. KIDS COUNT Special Report*. Annie E. Casey Foundation.

Fiszbein, A., & Schady, N. R. (2009). *Conditional cash transfers: reducing present and future poverty*. The World Bank.

Fryer, R. G., Jr., & Levitt, S. D. (2006). The black-white test score gap through third grade. *American Law and Economics Review*, *8*(2), 249–281. https://doi.org/10.1093/aler/ahl003

Gormley, W. T., & Gayer, T. (2005). Promoting school readiness in Oklahoma: An evaluation of Tulsa's pre-K program. *Journal of Human Resources*, *40*, 533–558.

Handa, S., & Davis, B. (2006). The experience of conditional cash transfers in Latin America and the Caribbean. *Development Policy Review*, *24*(5), 513–536. https://doi.org/10.1111/j.1467-7679.2006.00345.x

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, *24*(3), 411–482. https://doi.org/10.1086/504455

Hernandez, D. J. (2011). *Double Jeopardy: How third-grade reading skills and poverty influence high school graduation*. Annie E. Casey Foundation.

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. https://doi.org/10.1093/pan/mpl013

Hoddinott, J., Maluccio, J. A., Behrman, J. R., Flores, R., & Martorell, R. (2008). Effect of a nutrition intervention during early childhood on economic productivity in Guatemalan adults. *The Lancet*, *371*(9610), 411–416. https://doi.org/10.1016/S0140-6736(08)60205-6

Hu. (2016, November 22). *Plan boosts poverty alleviation efforts*. http://english.www.gov.cn/policies/policy_watch/2016/11/22/content_281475497354916.htm

Hungi, N., & Ngware, M. (2018). The effects of preschool participation on mathematics achievement of Grade 6 pupils in Uganda. *Educational Research for Policy and Practice*, *17*(2), 105–126..

Jenkins, J. M., Farkas, G., Duncan, G. J., Burchinal, M., & Vandell, D. L. (2016). Head Start at ages 3 and 4 versus Head Start followed by state pre-k: Which is more effective? *Educational Evaluation and Policy Analysis*, *38*(1), 88–112.

Jenkins, J. M., Watts, T. W., Magnuson, K., Gershoff, E. T., Clements, D. H., Sarama, J., & Duncan, G. J. (2018). Do high-quality kindergarten and first-grade classrooms mitigate preschool fadeout? *Journal of Research on Educational Effectiveness*, *11*(3), 339–374.

Jordan, N. C., Glutting, J., Ramineni, C., & Watkins, M. W. (2010). Validating a number sense screening tool for use in kindergarten and first grade: Prediction of mathematics proficiency in third grade. *School Psychology Review*, *39*(2), 181–195. https://doi.org/10.1080/02796015.2010.12087772

King, G., & Zeng, L. (2007). When can history be our guide? The pitfalls of counterfactual inference. *International Studies Quarterly*, *51*(1), 183–210.

Lesnick, J., Goerge, R., Smithgall, C., & Gwynne, J. (2010). *Reading on grade level in third grade: How is it related to high school performance and college enrollment*. Chapin Hall at the University of Chicago.

Li, W. (2017). *Li Wei: Popularize early childhood education in rural China and promote children's development. Keynote lecture at the Symposium of Early Childhood Education in Rural China*. http://www.qnzk.org/DocView.aspx?chnid=2&docid=1690

Lipsey, M. W., Farran, D. C., & Hofer, K. G. (2015). A randomized control trial of a statewide voluntary prekindergarten program on children's skills and behaviors through third grade. Research report. *Peabody Research Institute*.

Lipsey, M. W., Hofer, K. G., Dong, N., Farran, D. C., & Bilbrey, C. (2013). Evaluation of the Tennessee voluntary prekindergarten program: Kindergarten and first grade follow-up results from the randomized control design. Research report. *Peabody Research Institute*.

Magnuson, K. A., Ruhm, C., & Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance? *The Economics of Early Childhood Education*, *26*, 33–51.

McCoy, D. C., Zuilkowski, S. S., Yoshikawa, H., & Fink, G. (2017). Early childhood care and education and school readiness in Zambia. *Journal of Research on Educational Effectiveness*, *10*(3), 482–506. https://doi.org/10.1080/19345747.2016.1250850

Miller, L. C., & Bassok, D. (2019). The effects of universal preschool on grade retention. *Education Finance and Policy*, *14*(2), 149–177. https://doi.org/10.1162/edfp_a_00248

Ministry of Education of China, Office of the Press Secretary (2017). *Y. Lv asserts that the gross enrollment rate of early childhood education in China will be 85% by 2020*. Xinhua Net. http://www.xinhuanet.com/politics/2017-09/28/c_1121741001.htm

Ministry of Education of China. (2019). *Statistical announcement on the implementation of national education funds in 2018*. http://www.moe.gov.cn/srcsite/A05/s3040/201910/t20191016_403859.html

Montgomery, J. M., Nyhan, B., & Torres, M. (2018). How conditioning on posttreatment variables can ruin your experiment and what to do about it. *American Journal of Political Science*, *62*(3), 760–775. https://doi.org/10.1111/ajps.12357

Nores, M., & Barnett, W. S. (2010). Benefits of early childhood interventions across the world: (Under)investing in the very young. *Economics of Education Review*, *29*(2), 271–282. https://doi.org/10.1016/j.econedurev.2009.09.001

Nota, L., Soresi, S., & Zimmerman, B. J. (2004). Self-regulation and academic achievement and resilience: A longitudinal study. *International Journal of Educational Research*, 41(3), 198–215. https://doi.org/10.1016/j.ijer.2005.07.001

Puma, M., Bell, S., Cook, R., & Heid, C. (2010). *Head Start impact study (Final report)*. U.S. Department of Health and Human Services, Administration for Children and Families.

Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F., Mashburn, A., & Downer, J. (2012). *Third grade follow-up to the head start impact study: Final report. OPRE Report 2012-45*. Administration for Children & Families.

Ramey, C. T., Campbell, F. A., Burchinal, M., Skinner, M. L., Gardner, D. M., & Ramey, S. L. (2000). Persistent effects of early childhood education on high-risk children and their mothers. *Applied Developmental Science*, 4, 2–14.

Reynolds, A. J., Temple, J. A., Ou, S.-R., Arteaga, I. A., & White, B. A. B. (2011). School-based early childhood education and age-28 well-being: Effects by timing, dosage, and subgroups. *Science (New York, N.Y.)*, 333(6040), 360–364. https://doi.org/10.1126/science.1203618

Richter, L. M., Daelmans, B., Lombardi, J., Heymann, J., Boo, F. L., Behrman, J. R., Lu, C., Lucas, J. E., Perez-Escamilla, R., Dua, T., Bhutta, Z. A., Stenberg, K., Gertler, P., Darmstadt, G. L., & Paper 3 Working Group and the Lancet Early Childhood Development Series Steering Committee (2017). Investing in the foundation of sustainable development: pathways to scale up for early childhood development. *Lancet (London, England)*, 389(10064), 103–118. https://doi.org/10.1016/S0140-6736(16)31698-1

Rosenbaum, P. R. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society. Series A (General)*, 147(5), 656–666. https://doi.org/10.2307/2981697

Schweinhart, L. J. (2016). Lessons on sustaining early gain from a life-course longitudinal study. In A. J. Reynolds & A. J. Rolnick (Eds.), *Sustaining early childhood gains*. Cambridge University Press.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.

Skibbe, L. E., Connor, C. M., Morrison, F. J., & Jewkes, A. M. (2011). Schooling effects on preschoolers: Self-regulation, early literacy, and language growth. *Early Childhood Research Quarterly*, 26(1), 42–49.

Snow, C. E. (2006). What counts as literacy in early childhood? In *Blackwell handbook of early childhood development*. Blackwell.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. National Academies Press.

Tarullo, L. B., Xue, Y., & Burchinal, M. R. (2013, April). Are two years better than one? Examining dosage of Head Start attendees using propensity score matching methodology [Paper presentation]. The Biennial Meeting of the Society for Research in Child Development, Seattle, WA.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Burgess, S., & Hecht, S. (1997). Contributions of phonological awareness and rapid automatic naming ability to the growth of word-reading skills in second-to fifth-grade children. *Scientific Studies of Reading*, 1(2), 161–185. https://doi.org/10.1207/s1532799xssr0102_4

UNICEF (2015). Early learning, Provision of childcare and preschool education, A vision for early childhood. Retrieved March 19, 2016, from http://www.unicef.org.tr/sayfa.aspx?id=19&dil=en&d=1

Vandell, D. L., Belsky, J., Burchinal, M., Steinberg, L., Vandergrift, N., & NICHD Early Child Care Research Network. (2010). Do effects of early child care extend to age 15 years? Results from the NICHD study of early child care and youth development. *Child Development*, 81(3), 737–756. https://doi.org/10.1111/j.1467-8624.2010.01431.x

Vernon-Feagans, L., Garrett-Peters, P., Willoughby, M., Mills-Koonce, R., & Family Life Project Key Investigators. (2012). Chaos, poverty, and parenting: Predictors of early language development. *Early Childhood Research Quarterly*, 27(3), 339–351.

Walker, S. P., Wachs, T. D., Grantham-McGregor, S., Black, M. M., Nelson, C. A., Huffman, S. L., Baker-Henningham, H., Chang, S. M., Hamadani, J. D., Lozoff, B., Gardner, J. M. M., Powell, C. A., Rahman, A., & Richter, L. (2011). Inequality in early childhood: Risk and protective factors for early child development. *The Lancet*, *378*(9799), 1325–1338. https://doi.org/10.1016/S0140-6736(11)60555-2

Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathemat- ics, language, literacy, executive function, and emotional skills. *Child Development*, *84*(6), 2112–2130. https://doi.org/10.1111/cdev.12099

Wen, X., Leow, C., Hahs-Vaughn, D. L., Korfmacher, J., & Marcus, S. M. (2012). Are two years better than one year? A propensity score analysis of the impact of Head Start program duration on chil- dren's school performance in kindergarten. *Early Childhood Research Quarterly*, *27*(4), 684–694. https://doi.org/10.1016/j.ecresq.2011.07.006

Zaslow, M. J. (2013). *Investing in our future: The evidence base on preschool education*. Foundation for Child Development, Society for Research in Child Development.

Zhang, C. (2015). Equality and quality: To build nearby kindergarten to children in villages. *China Education News*. http://www.jyb.cn/china/gnxw/201503/t20150320_616552.html

Zhou, P. (2020, August 27). *China's Hukou system*. https://www.thoughtco.com/chinas-hukou-system-1434424