WFP EVALUATION

Impact Evaluation of the Preschool Nutrition Pilot in Selected Counties of Xiangxi Prefecture, Hunan, PR China

01/09/2018-19/11/2021

Decentralized Evaluation Report

DE/CHINA/2018/044
WFP China Office

November 2021

WFP

World Food Programme

SAVING LIVES CHANGING LIVES

Key Personnel for the Evaluation

WFP CHINA OFFICE

Evaluation Manager Ms. Han JIANG

PREPARED BY

Kevin CHEN, Team Leader and International Evaluator

Harold Alderman, International Evaluator

Chengfang LIU, Senior National Evaluator

Jieying BI, Intermediate National Evaluator

Haiquan XU, Intermediate National Evaluator

Yue ZHAN, Junior National Evaluator

Yanying YU, Junior National Evaluator

Acknowledgements

Disclaimer

Contents

List of Figures

List of Tables

Executive Summary

1.	This is the Endline Report for the impact evaluation of WFP's Preschool Nutrition Pilot Program in selected counties of Hunan province in Central South China, a project that aims to improve nutrition, cognitive, and socio-emotional conditions of preschool children and to increase smallholders' agricultural production and income through providing school meal subsidy, upgrading kitchen facilities, enhancing nutrition education of caregivers and preschool personnel and involving local poor smallholder farmers in the preschool food supply chain.

2.	Rapid economic growth over the past four decades has contributed to China's progress on eliminating early childhood malnutrition, concerns rise on prominent rural-urban disparities in early nutrition. The prevalence of anemia, stunting, and underweight among children under 5 were 5.38%, 1.12%, and 1.37% respectively in 2019, far below the set goals for the end of 2020 (12%, 7% and 5%, respectively) . However, children under 5 in rural areas are more vulnerable than their peers in urban areas. Most recent national statistics showed that the stunting rate for rural children was 5.8% by 2020, higher than the national average level of 1.12%. In addition, nutrition status of rural children, especially in poor areas, are more vulnerable than urban children when facing economic shocks and unexpected events .

3.	WFP China Country Strategic Plan (CSP) 2017-2021 commits to assist the country to achieve the national SDG target, which is reducing the stunting rate of children under 5 to below 7% and reducing anemia prevalence to less than 12% by 2020 . Hunan province is prioritized in the CSP, given the concentration of poverty and prevalence of malnutrition there. Since 1980s, WFP has implemented three assistance programs in 12 counties/districts in Hunan province, and played an important role in rural infrastructure improvement and poverty alleviation in western Hunan.

4.	By the time the nutrition pilot under discussion was introduced, as far as we know, there has been no national-level nutrition program in China targeting 3-5 years old. Although the government has targets for preschool enrolment, there is no platform for public food provision targeted at preschool aged children. Obviously, there is a gap to be filled. Therefore, WFP China Office seeks to fill this gap by introducing this nutrition pilot among preschoolers in Xiangxi Prefecture, Hunan Province.

5.	The impact evaluation of this nutrition pilot is commissioned by the World Food Program (WFP) China Office and covers the period from April/2018 to November /2021[1]. The purpose of this Endline Report is to: 1) overview the design and activities of the evaluation 2) document the fieldwork, data collection and management, and analyses of data from both the baseline and endline surveys; and 3) report key evaluation findings. The expected primary users for this endline report are the WFP China Office and its partners[2] in decision-making, notably related to program implementation.

6.	Since the way that WFP selected the kindergartens/preschools for the interventions is not random, an experimental design with complete randomization was not achieved. To evaluate impact in the most methodologically rigorous manner for the context, a quasi-experimental design was used. Specifically, the team estimated the program impact by Differences-in-Differences (DID) methods to correct for observable and unobservable biases. Therefore, baseline (pre-treatment, conducted in September 2018) and endline (post-treatment, conducted in May 2021) surveys were designed to collect quantitative data from both the treatment (with interventions) and comparison (without interventions) preschools. In each of the two round of surveys, trained enumerators administered the questionnaire to the kids' caregivers, teachers, preschool principals, kitchen staff, as well as smallholders. Registered nurses measured the height, weight, and hemoglobin levels for each kid. Trained examiners tested the cognitive level of sample children using internationally recognized scales in a standardized way. The team also collected qualitative data through desk

---

[1] The evaluation period was originally from April/2018 to April/2021. However, the endline survey was delayed from June/2020 to May/2021 due to the COVID-19 pandemic, resulting one year delay of the whole evaluation period.

[2] Ministry of Agriculture, Ministry of Education, provincial agriculture authorities and county governments, comprising representatives from the Department of Education, Health, and Agriculture, are partners in the design and implementation of WFP activities.

review and analysis of relevant literature and documents, as well as in-depth case studies and focus group discussions.

7.   The endline survey took place during the period of May 14-29 in 2021. The three modules of the endline survey include: 1) The child module aimed to trace and interview the same 1,334 children who participated in the baseline survey as well as their primary caregivers, teachers, principals, and kitchen managers to estimate the impact of the school feeding program on children's nutrition and development; 2) The smallholder farmer module contained a one-on-one household survey, together with in-depth focus group studies and key informants' interviews to explore the impact of the school feeding program on smallholder's agricultural production, marketing and income; 3) The cost module interviewed key informants of stakeholders from the WFP project offices and relevant agencies at the provincial, prefecture and county levels, as well as the 26 sample preschools with questionnaires to calculate the total cost and assess the cost-effectiveness of the school feeding program.

8.   The key findings of the evaluation team are summarised below, structured according to the evaluation criteria.

9.   Findings for Evaluation Question 1. What is the impact of nutrition program on nutrition and development of preschool children aged 3 and 5?

- **Nutrition:** Overall, we found no impact of the project on child nutrition outcomes, but with the caveat of potential spillover effects. The prevalence of anemia declined between baseline and endline. However, this decrease was similar in treatment and comparison groups. The project therefore had no impact on reducing anemia; The project also had no impact on anthropometric outcomes.

- **Cognition**: The analysis uncovered that the nutrition program had no impact on children's cognitive ability. However, results from heterogeneity analyses showed that the nutrition project does help improve the cognition of children from economically better-off families.

- **Social Emotions**: The school feeding program had positive effects on children's social emotions. It did help reduce the abnormal rates of hyperactivity of children on average. Besides, the program turned out to be effective in reducing the abnormal rate of total difficulties for some sub-groups of children. Specifically, the program significantly reduced the abnormal rates of hyperactivity and total difficulties in the younger children; The program helped reduce abnormal rates of emotional symptoms and hyperactivity for children from poor families.

- **Mechanisms**: We examined three underlying mechanisms for the above-mentioned results, including nutrition knowledge, feeding practice and interactive parenting. However, results show that the program did not have any significant impact on improving either the nutrition knowledge of primary caregivers, teachers or preschool kitchen workers, nor improving the feeding or parenting practices of primary caregivers. The findings help to explain why few significant impacts were found in the program evaluation process, pointing out the importance to focus on the behavior change strategies of targeted subgroups.

10.   Findings for Evaluation question 2. What is the impact of nutrition program on local smallholders?

- Both good practices and challenges of the direct purchase model implemented in Xiangxi were also uncovered. We find that direct purchase from smallholder farmers contributed to local smallholders' vegetable production diversity, income, nutrition knowledge and market engagement. In addition, based on results from analyzing the 106 households we interviewed in the endline survey, we find that smallholders in the treatment group had a higher dietary diversity score (DDS) than those in the comparison group. This program also has the potential to assist in building a local food system that is resilient, sustainable and inclusive. Moreover, this report identifies several constraints to be overcome in order to improve the current pilot. These challenges include the unstable supply under the farm to preschool model, the insufficient supply-side intervention, the transportation constraints and the expected decrease of demand from the local preschools.

- It is worth mentioning that the targeted beneficiaries of the smallholders varied in endline survey from the baseline survey. According to the initial selection criteria of smallholders in 2018,

beneficiary households should meet the following qualifications: 1) registered poor households; 2) capacity (e.g. land, labor) to engage in agricultural production; 3) a child attending the preschool with which they are about to sign the contract; and 4) signed agreement on a voluntary basis.

- The aforementioned criteria imply that the treatment group smallholders included in the baseline survey have to withdraw from the program when their children or grandchildren graduated from preschools. This led to high turnover rates of the beneficiaries. As a result, we are not able to construct a panel data for program evaluation. The findings are thus mainly based on a comparison between the treatment and comparison smallholders after the program interventions.

11.   Findings for Evaluation question 3. How appropriate is the program implementation?

- The program is well aligned with China's national priorities, including nutrition, child development and poverty reduction policies.

- The program is closely related to UN and WFP policies, strategies, and guidances on nutrition, child development, school feeding, and poverty reduction.

- The design and implementation of the program are partially in line with gender policies. Gender balance is achieved in terms of planned and actual numbers of children who received free school meals. In addition, women are also the main beneficiary of income-increasing activity by providing home-grown food for preschools and employment in the kitchen of preschools. However, the program does not provide a detailed analysis of different nutrition needs and health issues facing different gender groups and their implications.

- Administrative costs account for 10% of the total incremental costs, while labor costs by the treatment schools account for 20%. Together they account for nearly one-third of the total incremental costs during the implementation[3]. It is worthwhile to compare the cost composition of similar WFP programs in similar contexts to see if there are rooms to make the intervention more cost-effective from the administration and labor perspectives.

- Cost-effectiveness ratios vary significantly among subgroups of beneficiaries.

12.   The evaluation findings raise a number of important issues. Overall, we find that the program's direct purchase from smallholder farmers contributed to local smallholders' vegetable production diversity, dietary diversity, income, nutrition knowledge and market engagement. Outcomes related to child nutrition and development portray a more mixed picture. Measures of child nutrition, cognition and social emotions showed improvements from the baseline survey to the endline survey. However, while the improvement on some of these indicators was greater in treatment groups than comparison groups, this difference was not statistically significant. A strict interpretation of these findings is that the project had no impact on nutrition and development of preschool children. The Program was evaluated against a backdrop of China's national momentum towards reducing child malnutrition and widespread improvements in key maternal and child health indicators. This was also one of the explanations that the project did not show impacts as child nutrition in both treatment and comparison groups are improving, possibly thanks to the national efforts.

13.   Recommendations

- Recommendation 1: Building on its experience and reputation, the WFP should extend the provision of interventions in the study sites.

- Recommendation 2: The WFP should consider targeting at vulnerable sub-groups to optimize the impact of the project.

- Recommendation 3: The WFP should pay particular attention to the rise of child overweight and obesity.

---

[3] It should be noted that "incremental costs" means the extra costs incurred in the treatment group compared to the comparison group. Incremental costs are not equal to the amount actually paid by WFP and other donators.

- Recommendation 4: The WFP needs to take into consideration nutrition knowledge of caregivers to improve children's diet at home in addition to the nutritional value of the food choices offered in schools.

- Recommendation 5: The WFP should support further research to understand reasons for the decline in child malnutrition.

- Recommendation 6: The WFP should strengthen supply-side interventions to support local smallholders.

- Recommendation 7: The WFP should invest in rigorous evaluation at both country office and field level.

14.   There are a number of limitations to the evaluation. 1) There is concern that some selected participants may not show up in the baseline and endline survey due to various reasons. And participants in the comparison group who are not covered by the nutrition program might show little willingness to cooperate during the field survey. There is additional concern about attrition as the endline survey was originally planned to be undertaken three years after the baseline survey posing a potential risk to the evaluation. Supported by representatives from county governments, the evaluation team communicated clearly with comparison participants and proceed with caution. 2) The endline survey was delayed for one year due to the COVID-19 pandemic, rising extra concern about attrition. Since initially it was scheduled for June 2020, two school years after baseline survey, children aged 3-4 years old were expected to stay in the previous preschools at the endline survey time. However, after a one-year delay, children aged 4-5 years old were graduated and left their previous preschools. In addition, the pre-test before the endline survey indicated that a small proportion of children had transferred to other preschools. These situations might hinder tracking the same preschoolers in the endline survey. To minimize the attrition, the team conducted a where-about survey, based on which an optimal tracking plan was made.

1. Introduction

1.1. EVALUATION FEATURES

1.   This Endline Report is for the impact evaluation of the Preschool Nutrition Pilot Program in selected counties of Hunan province in Central South China, a project that aims to improve nutrition, cognitive, and socio-emotional conditions of preschool children and to increase smallholders' agricultural production and income through providing school meal subsidy, upgrading kitchen facilities, enhancing nutrition education of caregivers and preschool personnel and involving local poor smallholder farmers in the preschool food supply chain.

2.   The impact evaluation is commissioned by the World Food Program (WFP) China Office and covers the period from April/2018 to November /2021[4].

3.   Evaluation objectives: Like all evaluations at WFP, this evaluation serves accountability and learning purposes.

- Accountability: The evaluation is to assess and report on the performance and impact of the Preschool Nutrition Pilot Program in Xiangxi Prefecture, Hunan Province;

- Learning: The evaluation is to explore the reasons behind certain results occurred or not, thus to draw lessons, derive good practices and experiences for learning. It provides evidence-based findings to inform operational and strategic decision-making. Findings will be actively disseminated and lessons will be incorporated into relevant lesson-sharing systems. Given the pilot nature of the intervention, a stronger emphasis is expected on the learning purpose.

4.   The evaluation is expected to:

- assess the impacts of the nutrition program on children's nutrition status and development outcomes (including cognitive and socio-emotional conditions), as well as on smallholders' agricultural production and income level;

- analyze the cost-effectiveness of the program; and

- provide evidence-based findings to inform operational and strategic decision-making of the potential scale-up or continuation of the program. WFP HQ may use this evaluation for wider organizational learning and accountability.

5.   The evaluation team is led by Dr. Kevin Chen, Senior Research Fellow and Head of East and Central Asia Office (ECAO), International Food Policy Research Institute (IFPRI). He also serves as a Qiushi Chair Professor at Chinese Academy for Rural Development of Zhejiang University. Other members of the evaluation team include Dr. Harold Alderman from IFPRI Headquarter; Dr. Chengfang Liu and Dr. Renfu Luo from School of Advanced Agriculture Sciences, China Center for Agricultural Policy (CCAP), Peking University; Drs. Fengying Nie and Jieying Bi from Agricultural Information Institute (AII) of Chinese Academy of Agricultural Sciences (CAAS); Dr. Yanzhi Guo and Dr. Haiquan Xu from the Institute of Food and Nutrition Development (IFND) of Ministry of Agriculture and Rural Affairs (MARA); Ms. Yue Zhan and Zimeiyi Wang, research assistants at East and Central Asia Office of IFPRI; Dr. Xinghua Liu and Ms. Yanying Yu from Zhejiang University-IFPRI Center for International Studies at Zhejiang University.

6.   Impact evaluation timeline: The evaluation commenced on April 17, 2018 upon signing of the contract with WFP and all tasks will be completed by Oct 30, 2021. Building on the schedule set out in the ToR and updated in the proposal and agreements, the team prepared an overall timeline for the evaluation, including key activities and deliverables and their duration/due date (Annex 2).

---

[4] The evaluation period was originally from April/2018 to April/2021. However, the endline survey was delayed from June/2020 to May/2021 due to the COVID-19 pandemic, resulting one year delay of the whole evaluation period.

1.2. CONTEXT

7.   Rapid economic growth over the past four decades has contributed to China's success in eliminating extreme poverty in 2020, but challenges remain in reducing relative poverty and multidimensional poverty. Through its 13th Five-Year Plan for Economic and Social Development (2016-2020), the Government of China is implementing a major national effort to eliminate extreme poverty and substantially reduce relative poverty. By the end of 2020, China completed the arduous task of eliminating extreme poverty, which was 10 years ahead of schedule to achieve the first target of the UN 2030 Agenda for Sustainable Development[5]. Nevertheless, poverty in the post-2020 era is worthy of attention for a number of reasons. According to the World Bank, more than 300 million people in China are estimated to live on less than US$ 5.50 per day (2011 PPP) – the typical poverty line in upper-middle income countries. What's more, many of people that were brought out of poverty are still considered to be vulnerable to shocks. Relative poverty and multi-dimensional poverty will become increasingly prominent, mainly represented in the unequal access to income, social services and public services, health and education. Addressing relative poverty with accelerating changes in social structure, reducing vulnerabilities and preventing vulnerable people from falling back into poverty, reducing inequalities, and promoting human capital accumulation, represent emerging challenges that China needs to deal with in the future.

8.   Agriculture, forestry, and fishing account for 7 percent of China's GDP[6]. China has been able to feed its own population - which, according to the 2018[7] data accounts for 18.3 percent of the global population - with less than 10 percent of the world's farmland[8]. It also contributes substantially to China's food production. The estimated national budget for agriculture for 2019 was USD220,215 million[9]. The investment in modern technology – machinery, fertilizer and irrigation – and policy changes have been credited as fundamental drivers of China's agricultural growth. The National Plan to Promote Modern Agriculture (2016–2020) aims to maintain grain production at 550 million tonnes (Mt) and improve capacity to coordinate domestic and foreign markets[10]. It highlights the importance of appropriate import policies and structures for diverse products, channels and methods, and promotes international markets to complement China's comparative advantages. Recent policies place stronger emphasis on sustainable agricultural development in recognition of increasing global demographic, environmental and health challenges.

9.   China's agricultural sector is still largely dominated by rural smallholder farmers which, in 2016, totalled approximately 98.1 percent of China's 207 million registered agricultural producers - almost 50 percent of global smallholder farmers[11]. China's smallholder farmers face significant risks. Smallholder farming operations are less supported by national policies compared to their large-scale counterparts, and smallholder farms often do not have sustainable access to agricultural technologies to increase production, major markets to increase customer base, and food safety standard knowledge[12]. The movement of young, healthy people from rural areas to cities has meant that smallholder farmers tend to be ageing, female, and less educated.

10.   Although China has progressed on eliminating early childhood malnutrition, concerns rise on prominent rural-urban disparities in early nutrition. The prevalence of anemia, stunting, and underweight among children under 5 were 5.38%, 1.12%, and 1.37% respectively in 2019, far below the set goals for the end of 2020 (12%, 7% and 5%, respectively)[13]. However, children under 5 in rural areas are more vulnerable than their peers in urban areas. Most recent national statistics showed that the stunting rate for rural children

---

[5] Poverty Alleviation: China's Experience and Contribution. Retrieved at: http://www.gov.cn/zhengce/2021-04/06/content_5597952.htm

[6] World Bank https://data.worldbank.org/country/china?view=chart

[7] World Bank Open Data, 2018

[8] Report on China's Implementation of the Millennium Development Goals (2000-2015) July 2015

[9] https://www.oecd-ilibrary.org/sites/049d4bd3-en/index.html?itemId=/content/component/049d4bd3-en

[10] National Plan to Promote Modern Agriculture (2016–2020)

[11] UN People's Republic of China: Common Country Analysis 2020

[12] Ibid

[13] Statistical monitoring report for the Outline of the Development of Chinese Children (2011-2020). Retrieved at: http://www.stats.gov.cn/tjsj/zxfb/202012/t20201218_1810128.html

under age 6 was 5.8% by 2020, higher than the national average level of 3.5%[14]. Prevalence of underweight is also higher for rural children (2.4%), compared to their peers in urban areas (1.5%)[15]. In addition, nutrition status of rural children, especially in poor areas, are more vulnerable than urban children when facing economic shocks and unexpected events. Gender gaps in child nutrition are gradually improving. Most recent national statistics showed that on average, the prevalence of undernutrition for girls was lower than those for boys[16]: The prevalence of stunting was 4.2% for girls and 5.4% for boys; The prevalence of underweight was 1.9% for girls and 2.1% for boys; The prevalence of anemia was 1.9% for girls and 2.1% for boys.

11.   Being fully aware of the importance of fight against undernutrition among children, in the National Nutrition Plan (2017-2030) and China Child Development Outline (2021-2030), China sets the goal of bringing the anemia rate and the prevalence of stunting among children under five years old down to 10% and 5% by 2030, respectively.

12.   To achieve the goals, China has initiated a number of national programs (listed below) to promote child nutrition across different age groups, including children/students aged 0.5-2 and 6-15.

- China launched the School Milk Project in 2000 to ensure that students (6-15 years old) get safe, nutritious, and affordable dairy products in schools through the support of governments at all levels. By 2020, the project has covered more than 26 million students and 63 thousand schools across 31 provinces[17].

- The Chinese Ministry of Education (MoE) and 14 other departments jointly launched the Nutrition Improvement Program for Rural Students at the Compulsory Education Level (6-15 years old) (China's national school meal programme) in 2011 to address malnutrition among rural students by providing subsidized meals. By 2021, the program has covered 1552 counties in 28 provinces, benefitting about 38 million children every year[18].

- Nutrition Improvement for Children in Poor Areas program was initiated by the National Health Commission (NHC) in 2012, to promote soybean-based and micronutrient-fortified sachets, an in-home complementary food supplement, to improve the nutritional status of children aged 6-24 months in poor areas.

13.   By the time the nutrition pilot under discussion was introduced, as far as we know, there has been no national-level nutrition program in China targeting 3-5 years old. Although the government has targets for preschool enrolment, there is no platform for public food provision targeted at preschool aged children. Obviously, there is a gap to be filled. Therefore, WFP China Office seeks to fill this gap by introducing this nutrition pilot among preschoolers in Xiangxi Prefecture, Hunan Province.

14.   WFP China Country Strategic Plan (CSP) 2017-2021 commits to assist the country to achieve the national SDG target, which is reducing the stunting rate of children under 5 to below 7% and reducing anemia prevalence to less than 12% by 2020[19]. Hunan province is prioritized in the CSP, given the concentration of poverty and prevalence of malnutrition there. Since 1980s, WFP has implemented three assistance programs in 12 counties/districts in Hunan province, and played an important role in rural infrastructure improvement and poverty alleviation in western Hunan.

15.   Hunan also attracts the attention of other international agencies. Since 2012, it has implemented a program supported by $93.2 million loan from IFAD to improve rural community infrastructure and support sustainable agricultural development and marketing in the poorest, least fertile, least accessible and least developed areas of the province, including those in Wuling mountainous area. World Bank is demonstrating an approach to manage heavy metal pollution and improve agricultural land quality in Xiangxi and lend $100 million from 2017 to 2023.

---

[14] Report on Nutrition Development and Chronic Disease in China (2020)

[15] Ibid

[16] Ibid

[17] Promotion Plan of National Drinking Milk Program for Students (2021-2025).

[18] To Raise the Basic National Standards for Subsidizing Nutritious School Meals, 2021, MoE.

[19] WFP China Country Strategic Plan 2017-2021, 2017, World Food Programme.

1.3. SUBJECT BEING EVALUATED

16.　　Subject of the evaluation**:** Preschool Nutrition Pilot Program in Selected Counties of Hunan, China

17.　　Geographic scope: The program is conducted in two counties of Xiangxi Tujia and Miao Autonomous Prefecture, Hunan province of China, namely Longshan county and Yongshun county (please refer to the map in Annex 1). Both of them were nationally designated poverty counties in 2018. Specifically, benefited preschools were located in 13 townships, including Luota, Hongyanxi, Liye, Dianfang, Xiluo and Shipai in Longshan county, and Wanmin, Duishan, Xiaoxi, Yongmao, Huilong, Furong and Runya in Yongshun county.

18.　　Program outputs:

- Children in 25 selected preschools to be provided with nutritious school meals;

- Preschools to receive kitchen facilities and equipment;

- Nutrition knowledge of children, caregivers, preschool principals, teachers and kitchen staff to be improved; and caregivers' feeding practices and interactive parenting for children to be improved;

- At least 30% of food provided to the preschools to be sourced from local smallholder farmers.

19.　　Program outcomes in design:

- Improved nutrition (increasing height and weight; declining anemia rate) among preschoolers;

- Improved cognitive and socio-emotional conditions among preschoolers; and

- Increased smallholder farmers' income and agricultural production.

20.　　Key activities:

- Subsidize 4 Chinese Yuan per preschooler per school day to improve school meals in a total of 33 preschools which were not covered by any national nutrition improvement program. The enrollment of those beneficiary preschools ranges from 1,559 to 1,935 children, with an average of 1805. A conservative estimate of the covering children is above 3,000.

- Upgrade kitchen facilities and dining environment in the preschools;

- Enhance nutrition education through various activities and awareness campaigns, targeting at children, caregivers, preschool teachers, principals, and kitchen staff;

- Procure agricultural products from local poor smallholder farmers, connecting them with selected preschools. Trainings on farming techniques and food safety will be organized.

21.　　Main partners:

- Government bodies at province, county, and township levels, including agriculture, education, and health authorities, are partners in the design and implementation of the program.

- Private partners channel resources to support the program. As the complementary of the free meals, Mengniu Dairy Group provides each child a box of milk per day. Meituan.com, a group buying website, invites chief cook from Michelin-star-level restaurants to develop and share with preschools the localized menu, accompanied by nutrition education activities twice a year.

- China Development Research Foundation (CDRF) provides monitoring support for the program using their digital technology and monitoring platform. Kitchen manager in each benefited preschool uploads key information of daily meals, such as menu and procurement data, to the platform.

- Academic institutions, such as Chinese Center for Disease Control and Prevention and Institute of Food and Nutrition Development, are invited by WFP to provide nutrition trainings to caregivers, preschool principals and teachers.

22.　　Resources and budget: The designed programme budget was CNY 5,000,000 while the actual expenditure at the end of the project was CNY 4,231,844. Among the total cost, CNY 3,134,944 was used for

improving the quality of preschool meals, CNY 524,095 for conducting nutrition education, CNY 227,214 for upgrating kitchen equipment, CNY 27,965 for supporting local smallholders, and the rest CNY 317,626 for programe management. WFP China also succeeded in mobilising CNY 150,000 from Institute of Food and Nutrition Development of MARA for enhancing nutrition education. WFP China also contracted with China Development Research Foundation (CDRF) with additional CNY 222,560 for digital monitoring of the meals.

23. The evaluation works include (1) conducting a baseline survey on the nutritional, cognitive and socioemotional status and development of the current preschool children with special lens on left-behind children and girls. (2) conducting an impact evaluation (at terminal stage) to examine the overall effectiveness and impact of the preschool nutrition intervention program in terms of children's nutrition improvement, child development as well as linkage with smallholder farmers' production and promotion to local economic development.

24. Theory of change. The theory of change of this program is attached in Annex 3.

1.4. EVALUATION METHODOLOGY, LIMITATIONS AND ETHICAL CONSIDERATIONS

1.4.1. SAMPLING FRAMEWORK

25. The sample size of this study was beyond the control of evaluation team given that treatment schools were preselected. Therefore we chose a sub-optimal method. We first enrolled all treated preshools whose number of 3 and 5 years old children were more than 20 within two study counties. As a result, 13 preschools were selected as treatment preschools. And another 13 comparison preschools were chosen by the evaluation team based on the following criteria. The evaluation team obtained a list of all preschools from each of the two sample counties, with the enrollment information by age and gender. Then, the team took a 2-step approach to select samples for the purpose of the evaluation. First, the team excluded preschools (including those on the list of treatment group) whose number of 3 and 5 years old children were less than 20. Second, for each preschool in the treatment group, the team selected an most comparable untreated preschool in terms of the number and gender composition of 3 and 5 years old children from the same county to serve as its comparison.

26. Statistical power calculations were conducted using clustersampsi in Stata 15. According to our sampling strategy, 50 kids 3 or 5 years of age were sampled in each preschool (cluster). Data from the baseline survey of the project show that the intracluster correlation (ICC) for the Working Memory Index is 0.008. To detect an effect size of 0.20 standard deviations with 80% power and 95% significance level, when we run the code "clustersampsi, mu1(0) mu2(.2) alpha(.05) beta(.8) rho(0.008) m(50)" in Stata, we find that 12 clusters are required per experimental arm. For the stunting prevalence, the ICC is 0.004. When we run the code "clustersampsi, mu1(0) mu2(.2) alpha(.05) beta(.8) rho(0.004) m(50)" in Stata, we find that 11 clusters are required per experimental arm.

27. Among the 1,334 sample children in the selected 26 preschools, 600 of them were from Longshan county (44.98%), 734 were from Yongshun county (55.02%). 637 belonged to the treatment group (47.75%), 697 belonged to the comparison group (52.25%). 542 of them (40.63%) were 3 years old, 630 were 5 years old (47.23%). 645 are boys (48.35%), and 689 are girls (51.65%).

28. In the baseline survey, 146 households were interviewed. Of which, 72 households were from the treatment group whereas the rest 74 from the comparison group. In the endline survey, 106 smallholders were interviewed. Of which, 58 belonged to the treatment group whereas the rest 48 the comparison group. The 58 treatment group smallholders were those who were supplying food to preschools at the time interviewed. It is interesting to note that 12 of the 58 smallholders in the treatment group were those supplying to preschools at the time when the baseline survey was conducted. The comparison group are comprised of those smallholders in the baseline survey whom we managed to track during the endline survey (27 households). In addition, 21 households were newly included in the endline survey. They are from caregivers of preschools in the comparison group which did not receive any intervention from this nutrition program.

1.4.2. EVALUATION QUESTIONS

29. Evaluation Questions (EQs): The team developed three EQs and a set of sub-questions guided by the questions specified in the ToR and the Theory of Change analysis (Annex 3). Details around the EQs are set out in a full evaluation matrix (Annex 4).

- Evaluation Question 1. What is the impact of nutrition program on nutrition and development of preschool children aged 3 and 5[20]?

    Sub-question 1. To what extent does the nutrition program improve preschoolers' nutrition status?

    Sub-question 2. To what extent does the nutrition program improve preschoolers' cognitive ability?

    Sub-question 3. To what extent does the nutrition program improve preschoolers' socio-emotional status?

    Sub-question 4. To what extent did the nutrition program improve the nutrition knowledge of preschool personnel and caregivers?

    Sub-question 5. To what extent did the nutrition program improve the food consumption and nutrient intake of children?

    Sub-question 6. To what extent did the nutrition program improve the interactive parenting of caregivers?

- Evaluation question 2. What is the impact of the nutrition program on local smallholders?

    Sub-question 1. To what extent does the nutrition program boost agricultural production of smallholders?

    Sub-question 2. What is the impact of nutrition program on production diversification of smallholders?

    Sub-question 3. To what extent does the nutrition program contribute to local smallholders' income?

    Sub-question 4. What is the impact of the nutrition program on dietary diversity of smallholders?

- Evaluation question 3: How appropriate is the program implementation?

    Sub-question 1. To what extent are the program's designed objectives, targeting strategies, and activities consistent with relevant national policies, strategies, and guidance on nutrition, child development, and poverty reduction?

    Sub-question 2. To what extent are the program's designed objectives, targeting strategies, and activities consistent with UN and WFP policies, strategies, and guidance on nutrition, child development, school feeding, and poverty reduction?

    Sub-question 3. Is the program designed and implemented based on a sound gender analysis that is in line with national, UN, and WFP gender policies, considering the distinct needs and participation of boys/males and girls/females?

    Sub-question 4. Is the program cost-effective?

1.4.3. DATA COLLECTION

30. The data collection utilized questionnaires administrated by trained enumerators. The survey team developed these questionnaires for the primary caregivers of the sample preschoolers, their teachers, their preschool principals, as well as their preschool canteen staff. In the meantime, physical examination and cognition test were administered to preschoolers on a one-on-one basis. Secondary data from local governments, program offices, and other sources are collected as well. In addition, the evaluation team also conducted desk review and analysis of relevant literature and documents, as well as case study during field missions. Conclusions of the evaluation are drawn objectively from findings based on evidence. Gender lens were applied throughout the evaluation and integrated into key deliverables.

---

[20] The findings for sub-question 4-6 are presented in the mechanism analysis module to discover the underlying mechanisms behind the impacts of program on children's nutrition status, cognitive ability and social emotions, which are the findings for sub-question 1-3.

31.   Questionnaire survey: To collect data on indicators reflecting the impact, effectiveness, and efficiency of the program, the evaluation team developed a set of questionnaires for preschool children, their caregivers, preschool principals, teachers, kitchen managers, and smallholders, plus questionnaires to collect information about the costs of the nutrition program at WFP-, government-, preschool-, household-, private sector-, and nongovernment organization-level to calculate the cost-effectiveness of the program. The questionnaires include specific gender questions that enable disaggregation of the data to explore whether the impact varies by the gender of preschool children, caregivers, preschool personnel, and smallholders. According to the evaluation approach, a baseline field survey was undertaken during September 5 – 20, 2018, followed by an endline field survey in May 14-29, 2021 to capture the potential changes in key indicators before and after the intervention. Before the baseline and endline survey, a number of pre-test surveys had been organized to improve the survey forms and survey protocols (Details of the questionnaire surveys are attached in Annex 5 and fieldwork agenda can be found in Annex 6).

32.   Document review: The review relies on a range of relevant and available literature and documents from online resources and key stakeholders, especially the WFP and governments.

- During the inception phase, 80 documents were identified, located, filed electronically, and summarized in the literature review, providing background information on similar preschool nutrition programs worldwide and particularly in China, their impacts on children and smallholders, evaluation approaches and data analysis methods, and the cost-effectiveness of such programs. Literature review allowed the refining of the program's theory of change, improved the evaluation design, and provided clues toward understanding potential challenges. These background documents have been complemented by others as the evaluation proceeds. During the implementation and endline phase, literature shed light on monitor&evaluation of nutrition-sensitive programs were summarized, providing theoretical basis for tracking program performances and estimating program effects. Specifically, based on the literature, the team understood quality risks attributed to attrition, spillover, leakage and substitution and found corresponding solutions.

- The evaluation team expected to cooperate with stakeholders participating in/benefited from the nutrition program for relevant documentation of 1) policies, strategies, and guidance at various levels of governments, as well as UN system and the WFP, to analyze the consistence between the program design and policy environment. Particularly, gender policies of UN, WFP, and national governments have been closely followed; 2) program documentations about the details of program operation, management, achievement, and supporting materials such as slides and guidebooks for nutrition/agricultural training to understand the program implementation comprehensively. It also can be means to provide triangulation on some of the quantitative data gathered; and 3) administrative data/information from governments, WFP, and relevant academic institutions (particularly, monitoring data from China Development Research Foundation) to provide information and benchmarks for evaluation indicators and cross-check with primary data collected during the field missions.

33.   Case study: Case study was utilized for evaluating the nutrition program's impacts on smallholders. Rich information collected in case studies not only helped the evaluation team gain comprehensive understanding about the operational models and good practices of the nutrition pilot program, but also pinned down major challenges for smallholders to supply agricultural products to preschools, such as the diversified food items, demanding quality and food safety requirements, and high transaction cost. Semi-structured interview protocols were developed and used for the case studies.

34.   Key informant interview: In the inception phase, the evaluation team organized qualitative, in-depth interviews of key informants from governments, WFP and other UN agencies, and relevant academic institutions, to seek their first-hand knowledge about the socio-economic and policy backgrounds, program design, implementation, and monitoring. The interviews were meant to establish context, identify informed perspectives on program implementation, and triangulate the findings from quantitative methods. Interviews were loosely structured, leaving room to probe on relevant topics. In the smallholder farmer module, we also organized in-depth key informant interviews with those treatment households who participated both the baseline and endline surveys as well as local program officers to assess the impact of the program on agriculture production, market practices, income and nutrition. Last but not least, to collect cost data of the

program, we interviewed key informants of stakeholders from the WFP project offices and relevant agencies at the provincial, prefecture and county levels, as well as the 26 sample preschool principles.

35.   Focus group: Using focus group with a view to triangulation does more than simply make the data more complete. It also offers better understanding of the impact, effectiveness, and efficiency of the program. Focus group discussion with the participation of stakeholders along the foodstuff supply chain of the nutrition program was designed and organized in the endline survey to obtain information about the position of smallholders in the chain and the constraints/opportunities in their interaction with other participants such as input suppliers, traders, and preschool staff. The analysis of the supply chain helped indicate whether there was sufficient economic value in pursuing the nutrition program as an opportunity for smallholders and taking up additional efforts and associated costs to meet the quality, safety, and other standards prescribed by the program.

1.4.4. QUALITY ASSURANCE

36.   Limitations and risks are: 1) There is concern that some selected participants may not show up in the baseline and endline survey due to various reasons. And participants in the comparison group who are not covered by the nutrition program might show little willingness to cooperate during the field survey. There is additional concern about attrition as the endline survey was originally planned to be undertaken two years after the baseline survey posing a potential risk to the evaluation. Supported by representatives from county governments, the evaluation team communicated clearly with comparison participants and proceed with caution. 2) The endline survey was delayed for one year due to the COVID-19 pandemic, rising extra concern about attrition. Since initially it was scheduled for June 2020, two school years after baseline survey, children aged 3-4 years old were expected to stay in the previous preschools at the endline survey time. However, after a one-year delay, children aged 4-5 years old were graduated and left their previous preschools. In addition, the pre-test before the endline survey indicated that a small proportion of children had transferred to other preschools. These situations might hinder tracking the same preschoolers in the endline survey. To minimize the attrition, the team conducted a where-about survey, based on which optimal tracking plan was made.

37.   Quality assurance: Consideration was given to the following potential issues throughout the evaluation to ensure its impartiality, independence, credibility and utility.

- Spillovers[21]: To minimize the spillover effects, the team chose the comparison group carefully so that no two preschools were too close to each other to reduce the risk of contamination due to the spillovers of information or other treatment inputs. The team also included questions in the survey instruments to assess the risk that whether these types of spillovers occurred. Results supported that program might have caused spillover effects, given that school meal quality and nutrition awareness in the untreated preschools improved from baseline to endline. To identify and estimate spillover effects, control unites out of the program context, which are impossible to be affected by the direct/spillover effects of the program should be located and assigned. Nevertheless, the team were not allowed to follow this protocol due to limited time and budget. Fortunately, the spillover effects tended to be positive and lead to underestimated positive effects of the program.

- There might be substitution and leakage of the treatment among preschoolers in the treatment group. Specifically, substitution might happen as children in the treatment group were provided nutritious food in preschools, their families might provide less food or less nutritious food at home. Leakage might happen if the preschoolers in the treatment group take their ration of food from preschools to share with other family members at home. To capture whether these are an issue, the team asked the caregivers about children's food intake at home during the baseline and endline surveys. The results shown the diet quality of children in the treatment group had no significant differences compared to that of their counterparts, denying the existence of

---

[21] Spillover effects happen when an intervention affects a nonparticipant, and it might be positive or negative. However, we were not able to collect data in other untreated counties to quantify the spillover effects.

substitution. In addition, results failed to prove the existence of leakage given that few children (7.5%) in the treatment group brought food from preschools to home at the survey time.

- Impact Heterogeneity[22]: In fact, the effects of the program vary by characteristics at both the preschooler and their family levels, such as gender of the child and their socio-economics status. Therefore, the team estimated not only the overall effects of the nutrition program but also the heterogeneous ones.

- Hawthorne[23] and John Henry Effects[24]: Effects due to evaluation itself are a potential concern for the study given that blinding of study participants is impossible in this context. (This is true of any intervention that cannot be double-blinded.) We do not believe John Henry effects are a significant concern; however, it is possible that participants in the treatment groups change their behavior due to knowledge of participation in a novel government program (Hawthorne effect). While this may pose some challenges to external validity, the possibility of such effects is unfortunately unavoidable in this context.

- The evaluation team also made great efforts to ensure data used in the evaluation report were checked for accuracy and reliability. The evaluation report also acknowledges limitations to the generalize ability of the conclusions that could be drawn from the evidence.

- WFP has developed a Decentralized Evaluation Quality Assurance System (DEQAS) based on international good evaluation practices. It set out process maps and templates for evaluation products, including checklists for feedback on quality for each of the evaluation products. The principles and all appropriate content have been systematically applied to this evaluation.

1.4.5. ENDLINE SURVEY IMPLEMENTATION

Endline survey modules

38.	Just as we did in the baseline survey in 2018, the endline survey in 2021 includes three modules as follows.

39.	The child module aimed to trace and interview the same 1,334 children who participated in the baseline survey as well as their primary caregivers, teachers, principals, and kitchen managers in order to build a panel dataset for evaluating the program's impact on child's nutrition and development. The survey team conducted physical examination and cognitive tests for the sample children and collected the same contents of information from their primary caregivers, teachers, principals, and kitchen managers through one-on-one, face-to-face interviews undertaken by trained enumerators.

40.	The cost module interviewed key informants of stakeholders from the WFP project offices and relevant agencies at the provincial, prefecture and county levels, as well as the 26 sample preschools with questionnaires. Combined with cost documents provided by the stakeholders, the cost was calculated for assessing the cost-effectiveness of the school feeding program.

41.	The smallholder farmer module contained a one-on-one household survey, together with in-depth focus group studies and key informants' interviews. The purpose of this module is to explore the impact of the school feeding program on smallholder's agricultural production, marketing and income.

Tracking the sample children

42.	The endline survey was initially scheduled for June, 2020. Unfortunately, due to the Covid-19 pandemic, it was delayed by nearly one year. As the time interval between the baseline and endline surveys had been expanded, the proportion of children who graduated from preschools or transferred to other preschools was sizable, making it difficult to trace all children from the baseline. As it was critical to track as

---

[22] Impact heterogeneity refers to differences in impact due by beneficiary type and context.

[23] The Hawthorne effect is a type of reactivity in which individuals modify an aspect of their behaviour in response to their awareness of being observed.

[24] The John Henry Effect refers to the tendency for people based in comparison group to perceive themselves at a disadvantage to the treatment group and work harder to overcome the perceived deficiency.

many children from the baseline survey as possible to ensure the power and validity of the impact evaluation, significant extra efforts were made by the survey team to track the children.

43.   To capture the where-about information of each sample child, the survey team visited each of the 26 sample preschools immediately before the endline survey. Figure 1 showed the initial sample framework including all the sample children needed to be tracked and their whereabouts in April 2021, one month before the formal endline survey.

44.   Data from the whereabout form show that, by the endline survey, 72 (5%) sample children had transferred to another preschool or entered primary school located outside the two sample counties. Therefore, due to the severely binding time and budget constraints, it was not feasible for the survey team to conduct physical examination or cognitive tests on these children nor face-to-face interviews with their primary caregivers.

45.   The survey team successfully traced 1,210 out of the 1,334 (or 90.7%) sample children. In other words, the attrition rate is 9.3%. Of the 1,210 children who participated in both the baseline and endline surveys, 331 children (25%) remained in the 26 sample preschools, 139 children (10%) transferred to other preschools within the two sample counties, and the rest of the 740 children (55%) either graduated from one of the sample preschools or entered a primary school located within the two sample counties. The primary schools they entered were typically nearby township primary schools or primary schools located in the county seat.

46.   The survey team was not able to obtain the whereabout of 52 (4%) sample children by the endline survey due to three major reasons as follows. Firstly, by the endline survey time, 2 sample preschools with 87 sample children had stopped running. The team not only tried to get their whereabout from their previous preschool teachers or principals, but also tried to use contact information collected in the baseline survey to contact their caregivers. However, 13 sample children from these 2 closed preschools could not be tracked. Secondly, 38 sample children who transferred to other preschools or entered primary schools did not update their contact information and could not be located either with their contact information from the baseline survey. Thirdly, it is sad that one sample child unexpectedly passed away before the endline survey.
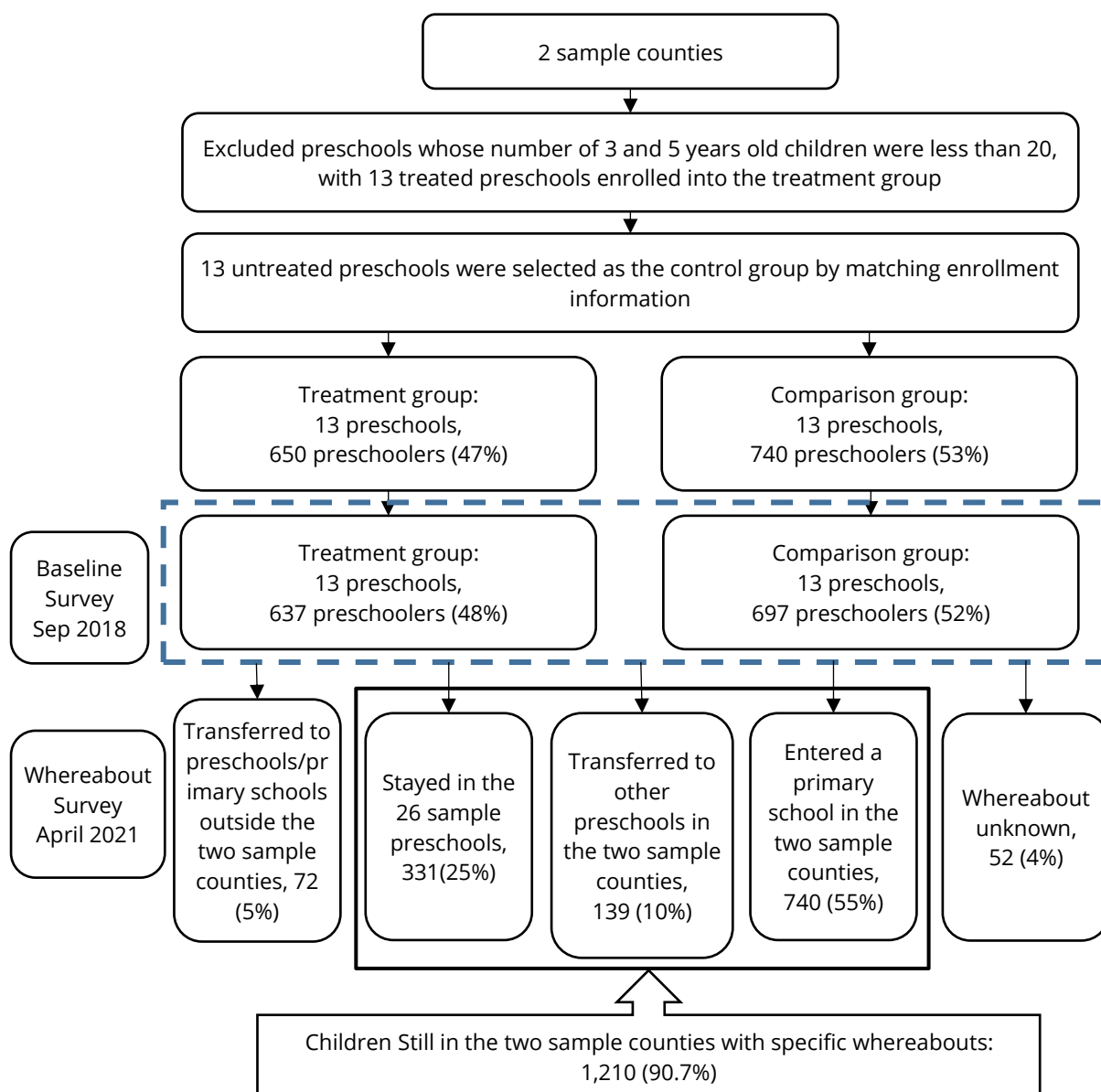
Figure 1 Sampling framework and whereabouts one month before the endline survey in May 2021

Training of enumerators and nurses

47.   The research team trained the enumerators and nurses intensively in a standardized way. In total, we recruited 76 enumerators from universities or research organizations, plus 10 registered nurses from local maternal and child care centers. These 86 people first gathered in Longshan county for a three-day intensive technical and ethical training, plus a one-day field practice run in nearby rural preschools.

48.   The training of enumerators who were responsible for child, smallholder farmers and cost survey was overseen by Dr. Kevin Chen from the IFPRI and Zhejian University and conducted jointly by Dr. Chengfang Liu from Peking University, Dr. Jieying Bi from Agricultural Information Institute (AII) of Chinese Academy of Agricultural Sciences (CAAS), Dr. Hanquan Xu from the Institute of Food and Nutrition Development (IFND) under the Ministry of Agricultural and Rural Affairs (MARA) and Dr. Xinghua Liu from Zhejiang University.

49.   Enumerator training was organized in two parts. The first part was focused on the questionnaires, module by module, question by question to make sure that the enumerators understand the questions in a standardized way. After they got familiar with the questionnaires, enumerators were trained to use tablets to complete the electronic questionnaires. Dr. Chengfang Liu, as a licensed trainer of the Chinese version of the fourth edition of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-IV) and the Wechsler

Intelligence Scale for Children (WISC-IV), administered the technical training of the cognitive test. 30 Qualified enumerators who passed the qualification test by Dr. Liu acted as cognitive test examiners in the field. Communication skills and ethical considerations were on the training agenda as well to ensure a smooth, safe, and respectful survey procedure. Followed the indoor training was the field practice run in a nearby rural preschool, during which questionnaire enumerators, cognitive test examiners got tested to ensure they have understood what they had learned in training.

50.   In the meantime, Dr. Renfu Luo from Peking University trained 10 registered nurses on how to take anthropometric measurements (height and weight) and hemoglobin tests for the sample children. The nurses also joined the field practice run to get familiar with the test equipment, test procedure, and ethical considerations.

Endline Data Collection

51.   Different from the baseline, we used electronic survey device in the endline survey. Specifically, Computer-Assisted Personal Interviewing (CAPI) was adopted in the caregiver and teacher module to replace paper-and-pen methods. This method relies on digital questionnaires to guide the interview and records data on a computerized device, which saves interview times and reduces errors compared to paper-and-pen methods. Tablets were used to conduct interviews and record respondent's answers during the field missions. To make sure that these electronic questionnaires helped to recognize outliers, avoid missing values and facilitate skip patterns for logical reasons, the team made multiple rounds of pretests and revisions. Moreover, a field pre-survey was also conducted during April 18-22 in 2021, one month before the formal endline survey to ensure the validity and stability of electronic questionnaires. Other modules such as smallholder farmer, cost, principal, kitchen manager were still collected using paper-and-pen method.

52.   The endline survey took place during the period of May 14-29 in 2021. 47 enumerators were sent to Yongshun County, while 31 to Longshan County to conduct the survey in preschools. The remaining 8 enumerators were sent to villages in both counties to conduct smallholder farmer surveys, first in Longshan and then in Yongshun.

53.   Ideally and as originally planned, the team should have returned to the same 26 sample preschools to conduct the endline survey in May 2020 to track and survey the same 1,334 children as well as their caregivers, teachers, principals and kitchen managers. Unfortunately, the endline was postponed by one year and part of sample children had transferred to other preschools or entered primary schools, the preschools/primary schools involved in the endline survey expanded from 26 to 131. As it turned out, it was not feasible for the survey team to visit all these 131 preschools/schools during the survey due to the binding manpower and budget constraints. Therefore, the team had to take a suboptimal approach. Specifically, the team visited 44 preschools/primary schools with relatively large number of sample children from the baseline survey. In the meantime, the survey team assembled sample children in the nearby preschools/schools as well as their caregivers to conduct the caregiver questionnaire, child's physical examination and child's cognitive test in those preschools/schools with relatively higher concentration of sample preschoolers. In this way, the survey team was able to track sample children as many as possible. There was cost associated with this way though. As we are concentrated in these preschools/schools, we did not have the time to visit the rest 87 preschools/schools, which have just one or two of our sample children per school.

54.   The data collection began on May 17, 2021 and lasted for 12 days. Enumerators, cognitive test examiners, and nurses were grouped to be responsible to undertake different modules of the endline survey while visiting the sample preschools/schools in groups. Enumerators responsible for smallholder farmer surveys visited selected samples in Longshan and Yongshun successively.

55.   Enumerators in preschools/schools started their work by briefing their interviewees about the purpose and contents of the survey. To protect the rights and welfare of preschool children, their caregivers were responsible for deciding whether the children would participate or not and complete the informed consent form. With interviewees being fully aware of the survey procedure and their own rights, the enumerators then completed required questionnaires for caregivers, principals, teachers, and kitchen staff. All the questionnaires, digital or printed, were carefully dated, completed, triple checked and stored by the end of the same day.

56.   Before the cognition test, a sample child was told that she/he was invited to participate in some interesting activities, and she/he should listen carefully and follow the directions of the examiners. All the

trained examiners were trained to help the child relax and approach the activities positively. All the test items were designed in such a way to be appealing and engaging to children.

57.   We focus on three health indicators: hemoglobin concentrations (Hb), height and weight. Hemoglobin levels were measured on-site using HemoCue Hb 301 systems. Height and weight measurements were also taken on site, following WHO standard protocol. The children were measured in light clothing without shoes, hats, or accessories. Weight was measured with a calibrated electronic scale recommended by scholars from the West China School of Public Health of Sichuan University. Body height was measured using a standard tape measure. The nursing team was trained to ensure that the weighing station was set up on level ground to ensure accuracy of the equipment. Two nurses manned each measurement station, with one responsible for preparing subjects for measurement (removing shoes, offering instructions, reassuring and positioning children, etc.) and the other responsible for conducting and recording the measurements.

58.   Table 1 shows the response rates by survey modules. In the child survey module, the evaluation team expected to collect data from 1,210 sample children and their caregivers. As mentioned, due to constrained manpower and time. We were not able to track 72 sample children who migrated with caregivers outside of the two sample counties, nor those 52 sample children for whom we did not have the updated contact of their caregivers, nor those 28 sample children who were too scattered in remote preschools/schools to be reached given the time, manpower and budget constraints. These situations combined, the expected follow-up sample size was 1,182 sample children who participated in the baseline survey. Well, there were a number of further complications though. Firstly, some caregivers were temporarily not at home or occupied with farm work or could not complete the questionnaire in person (usually due to mental/physical illness) on the survey day. Secondly, the sample children did not attend preschool/school on the survey day. Thirdly, a few sample children refused to take physical examination/cognitive test.

59.   In order to make up for these complications, the survey team conducted make-up surveys through phone calls. To do so, 8 enumerators who had participated in the whole endline survey and were familiar with the caregiver questionnaire were enrolled. After a rigorous training, they undertook phone interviews following standard protocols. The makeup phone survey began on September 20 and lasted for 7 days. We successfully tracked 42 caregivers.

60.   In sum, the survey team completed a total of 965 caregivers questionnaires, conducted physical examination for 1,053 children and child cognitive test for 1,004 children. The response rates for caregiver module, physical examination and cognitive test were 72%, 79% and 75%, respectively. The responses rates for other modules were 100%, with the only exception of the preschool principal and kitchen staff module. Comparatively speaking, the response rates were very close to that for similar early childhood programs under the Chinese contexts. For example, the response rates of the endline surveys for China Reach, an early childhood intervention program jointly conducted by China Development Research Foundation and Department of Maternal and Child Health Services, National Health and Family Planning Commission, were also around 70%-80%[25].

---

[25] The response rates were referred to the China Reach evaluation reports of Gansu program, Guizhou program and Xinjiang program.

Table 1 Final sample size in the endline survey

	Child Module					Cost Module	
	Caregiver	Child Physical Examination	Child Cognitive Test	Preschool Principal	Kitchen Staff	Program Management Office	Baseline Sample preschools
Baseline sample	1,334	1,334	1,334	26	26	—	—
Target follow-up sample	1,210	1,210	1,210	24	24	2	24
Follow-up	965[26]	1,053	1,004	24[27]	24	2	24
Response rates	72%	79%	75%	92%	92%	100%	100%

1.4.6. DATA ENTRY, CLEANING, AND MANAGEMENT

61.   All the printed questionnaires from the endline surveys were sent to professional data entry company, while data recorded in the digital questionnaires were downloaded from the electronic questionnaire headquarters. A team of 12 data analysts were organized to conduct data cleaning using STATA. The 2-week cleaning procedure included four steps: 1) Outliers, missing value and logical flaws were identified and recorded; 2) Based on the list of all the problematic data, the analyst then went back to cross-check the original questionnaires to see if they were mistakenly entered or if notes were used to explain the special cases ; 3) The remaining issues after questionnaire checking were solved through make-up telephone surveys (as mentioned earlier in this part); and 4) Correct data were reentered into the database and double-checked.

62.   Since benefited preschools and smallholders have been preselected by the WFP and local counties, a complete randomization will be impossible. Consequently, quasi-experimental approach was adopted. Specifically, the evaluation estimated the program impact by Difference-in-Differences (DID) methods. Details of methodology can be found in Annex 7.

1.4.7. COVID-19 LIMITATIONS ON PROJECT IMPLEMENTATION AND EVALUATION

63.   COVID-19 and the related policy measures to curb the spread of the virus are expected to have considerable impacts on the school feeding intervention and accompanying evaluation. China was the first country to be hit by the COVID-19 pandemic, where the first confirmed cases were reported in December 2019. Faced with the escalating virus outbreak, the Government implemented large-scale restrictions on mobility in late January 2020 in an attempt to limit COVID-19's spread. The main effect of COVID on project activities was that feeding activities were suspended from March to May 2020 as school closures were imposed nationwide. By the endline survey of the project, children in the treatment group had approximately two months less exposure to the project activities than was originally expected. This meant that during that time children lacked school meals that were directly contributing to their dietary intake and households' food security. Additionally, WFP suspended or canceled some nutrition education workshops, trainings, and meetings to avoid large gatherings during this period. However, local project office reported having moved some meetings and trainings online.

64.   Early reports and evaluations indicate that children who relied on nutrition services provided by schools may suffer from worsening health and nutritional status in the short and medium term. Nutrition shocks, especially for the youngest children, are also expected to have strong long-term impacts on test scores, educational attainment, income, absenteeism and health (Almond and Currie, 2011; Sudfeld et al., 2015; Andrabi, Daniels and Das, 2020). Furthermore, lost schooling and learning may also lead to poor nutrition and health for themselves and their children in the long term (UNICEF, 2021). The UN reports that

---

[26] In the field survey, the team had collected 921 valid caregiver questionnaires. Another 44 were collected during the supplementary telephone interviews.

[27] Two preschools had stopped running before the endline survey.

data show a deeply concerning trend: Since the beginning of the pandemic, UNESCO estimated that 1.6 billion learners in 199 countries worldwide were affected by school closures, with nearly 370 million children not receiving a school meal in 150 countries (UNESCO, 2020; WFP, 2020). In 2020, globally, an estimated 39 billion in-school meals have been missed during school closures. Children globally are estimated to have missed an average of 4 out of 10 in-school meals they would have regularly received, with children in some countries missing as many as 9 out of 10 in-school meals.

65.   That said, the restrictions and intensity of the pandemic experience did have an unexpected impact on parenting experiences. Migrant parents stayed at home longer after the Spring Festival due to travel restrictions and thus spent more time with their children. Teachers in our study sites reported that children were livelier during this period.

66.   It is also important to note the significant impact of COVID-19 on the evaluation process itself. Due to the budget constraint, the project did not carry out any evaluation activity at midline and limited documentation of participants' actions and food security and nutrition status during covid-19 should be acknowledged. Once Covid-19 related restrictions were lifted, and it was deemed sufficiently safe for research teams to travel to the field for research purposes, we managed to follow up the sample at the endline survey, which was postponed from June 2020 to May 2021. As data collection was largely conducted after the COVID-19 outbreak, respondents' attitudes, perceptions, and behaviors regarding project outcomes therefore have been affected by COVID-19. This means that it is possible that some potential effects of project activities could have been wearing off during this period.

67.   Longer than planned time lag between the baseline and endline surveys also presents a unique risk as it became more difficulty to trace the children who participated in the baseline survey. Specifically, children aged 4-5 in the baseline would have graduated from preschools into primary schools, children aged younger than 4 years may have transferred to other preschools outside of our sample preschools, or sample counties even sample prefecture. Although we tried our best to trace those children by visiting them in their current schools/preschools to conduct face-to-face interviews, or called their primary caregivers to conduct interview by telephone, we have to face the reality that we might miss data on some modules for some students, or some sample children have even missed. Either way, the data quality might be compromised during the survey postponement. We assessed the impact of COVID-19 on our sample to understand if it differentially impacted our treatment and comparison groups. Overall, we did not find differences in the impact of the Covid-19 pandemic on children from the treatment and comparison groups (detailed results can be found in Annex 8). This means that we expect Covid-19 would not affect the validity of our evaluation.

1.4.8. ETHICAL CONSIDERATIONS

68.   WFP decentralized evaluations must conform to WFP and UNEG ethical standards and norms. The contractors undertaking the evaluations are responsible for safeguarding and ensuring ethics at all stages of the evaluation cycle. This includes, but is not limited to, ensuring informed consent, protecting privacy, confidentiality and anonymity of participants, ensuring cultural sensitivity, respecting the autonomy of participants, ensuring fair recruitment of participants (including women and socially excluded groups) and ensuring that the evaluation results in no harm to participants or their communities.

2. Evaluation Findings

2.1. EVALUATION QUESTION 1: WHAT IS THE IMPACT OF NUTRITION PROGRAM ON NUTRITION AND DEVELOPMENT OF PRESCHOOL CHILDREN AGED 3 AND 5?

1.   The evaluation findings are first presented according to the key evaluation sub questions: children's nutrition status, cognitive ability, socio-emotional conditions. Accordingly, the evaluation methodologies required comparison between treatment group comparison groups in a range of outcome variables. In doing so, we also control for child and household characteristics.

2.   We thus first compare the baseline characteristics of the households in the treatment and comparison groups. These descriptions have two main objectives: (1) to provide context for the analyses of children's outcomes in subsequent sections, and (2) to assess if differences in health-related outcomes between the treatment and comparison groups might have been driven by differences in baseline characteristics rather than the impacts of the program.

3.   We presented evaluation results for each key outcome variable, where we also examined program impacts by subgroups, including those defined by socioeconomic status, children's gender, and age. For subgroup analyses, we restricted the sample to the relevant subgroup (for example, children from low-income households). In the following sections, we only report impacts for subgroups where the impact for that group is statistically significant. This approach enables us to focus on the most meaningful subgroup impacts.

4.   In this section we examined the demographic and socioeconomic characteristics of the households, primary caregivers, and children in treatment and comparison groups at the baseline survey. We first presented selected characteristics in Table 2 below, followed by a balance test of baseline outcome variables and a more complete set of characteristics in the Annex 9.

5.   A total of 1,334 children in 26 preschools were sampled at baseline. Most demographic and socioeconomic characteristics of the sampled children and their households were similar in the treatment and comparison groups. Two potentially important exceptions are ethnicity and income status, with children from ethnic minority and registered poor households being more likely to be from treatment schools. This is understandable as the program were supposed to be targeted at preschools with high proportions of children from ethnic minority and low-income households. From evaluation point of view, if ethnicity and/or income status are correlated with child outcomes, these differences at baseline could bias our impact estimates. To account for these differences, these two characteristics are always controlled for in adjusted regression models when estimating the impact of the program[28]. Given the similarity between treatment and comparison groups in most characteristics at the baseline, together with other control variables included in estimations, socioeconomic or demographic characteristics are unlikely to be driving the estimated impacts of the program on health-related outcomes of the preschoolers.

---

[28] As described in the methodology section, we also controlled for other characteristics—including father's education, child age, gender, left behind status—even though they were balanced in the treatment and comparison groups, to improve the precision of our impact estimates.

Table 2 Baseline socio-demographic characteristics of study sample

	Treatment Group	Comparison Group	Differences (1)-(2)	t-Test (p-value) H0: (1) = (2)
	(1)	(2)	(3)	(4)
Boy	0.53 [0.50]	0.50 [0.50]	0.03 [0.03]	0.37
Elder children	0.49 [0.50]	0.49 [0.50]	0.00 [0.10]	0.97
non-Han minority	0.93 [0.26]	0.85 [0.35]	0.07 [0.03]	0.02**
Left-behind children[29]	0.73 [0.44]	0.69 [0.46]	0.05 [0.03]	0.13
Registered poor family	0.42 [0.49]	0.27 [0.44]	0.15 [0.03]	0.00***
Subsistence allowance family	0.11 [0.31]	0.07 [0.26]	0.03 [0.02]	0.14
Longshan	0.42 [0.49]	0.47 [0.50]	-0.05 [0.11]	0.65
Father has at least senior high school degree	0.12 [0.32]	0.13 [0.34]	-0.01 [0.02]	0.50
Mother has at least senior high school degree	0.11 [0.31]	0.13 [0.34]	-0.02 [0.02]	0.32
Obs	637	697	1334	

Notes: a) Means with standard deviations reported in brackets. b) Cluster-robust standard errors adjusted for clustering at the class level in parentheses. c) *** p<0.01, ** p<0.05, * p<0.1

6.   To control for the socioeconomic status (SES) of the preschoolers, we created a wealth index by combining a range of household assets, following the approach of the Demographic and Health Surveys (DHS). The wealth index can be interpreted as a composite measure of a household's SES. This index enables us to rank and divide households into quintiles for subgroup analyses based on their relative wealth, with households in lower quintiles being less wealthy than those in higher quintiles. In terms of the distribution of the wealth index, as shown in Figure 2, the treatment group tends to be less wealthy. Again, this is somehow expected as the program was supposed to target at the preschools with high proportions of children from less wealthier households.

---

[29] Left-behind children refer to those children who are cared for by a single parent or other relatives, usually living in rural areas with one or both parents migrating to the cities for better jobs.
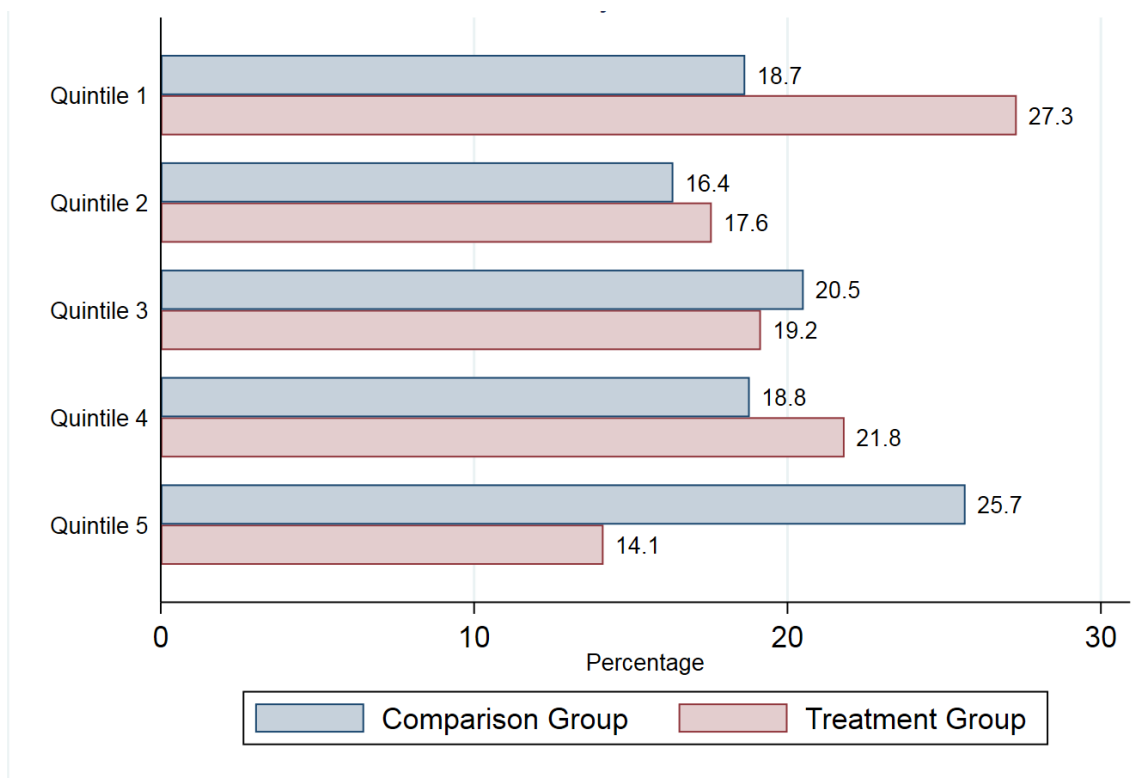
Figure 2 Wealth distribution, by treatment status

2.1.1. SUB-QUESTION 1: TO WHAT EXTENT DOES THE NUTRITION PROGRAM IMPROVE PRESCHOOLERS' NUTRITION STATUS?

7.   A key aim of the WFP Preschool Nutrition Pilot Program is to improve the nutritional status of beneficiary children. To assess nutritional status, both rounds of surveys included an anthropometric module in which all children were weighed, and their heights measured, and hemoglobin level tested to check whether a kid is anemic or not[30]. Using these measures, standard anthropometric indicators were calculated for children, and compared across treatment and comparison groups to isolate program impact on these outcomes from the baseline to the endline surveys.

Overall changes in child nutrition from baseline to endline

8.   Survey results show that at the baseline, anemia (33.36%) was the most prevalent malnutrition among sample children, followed by stunting (10.59%) and underweight (5.41%). According to the 2018 Surveillance Report of Outline for the Development of Children (2011-2020), the prevalence of anemia, stunting, and underweight among children under 5 were 5.44%, 1.11% and 1.43%, respectively, indicating that our sample children fell behind the national average. Our endline survey data show that there has been great progress in reducing child undernutrition in the project area. By the endline survey, the prevalence of anemia, stunting and underweight has declined to 6.88%, 5.44% and 4.38%, respectively. However, as shown in Figure 3, there's still room for improvement compared to national average in 2019.

---

[30] Anemia is defined as a hemoglobin concentration below a specified cutoff point, which for children under 5 years of age (<60 months) is 110 g/l at sea level and for children older than 5 years (≥60 months) is 115g/l at sea level.
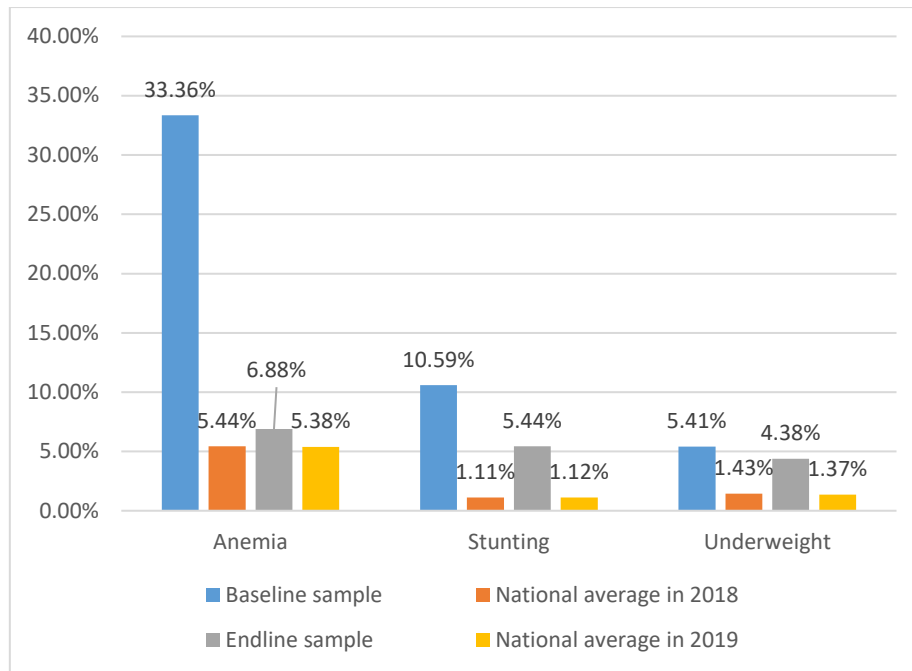
Figure 3 Prevalence of malnutrition in baseline and endline

9.   At the other end of the malnutrition spectrum, overweight (11.34%) and obesity (2.48%) existed in baseline but not a worrisome phenomenon. However, as shown in Figure 4, we saw a large increase for both indicators from the baseline to the endline surveys in our project area.
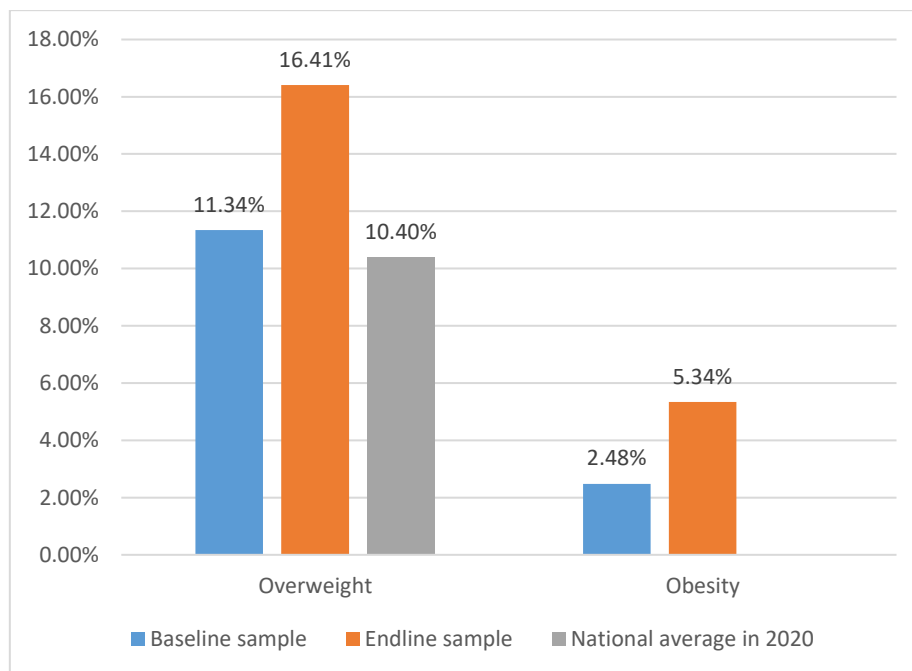
Figure 4 Prevalence of overweight and obesity by surveys

Program impact on the prevalence of anemia

10.   At the time of the baseline survey (2018), around 33% of children were anemic. Figure 5 below showed how this proportion changed in the treatment and comparison groups across the two survey rounds. The treatment group saw a large decline from the baseline to the endline surveys, from 32% to 6%. However, the prevalence of anemia decreased similarly in the comparison group during the same period, from 36% to 8%, indicating no impacts that could be attributable to the program of interest. Overall, results from

descriptive analyses show that the school feeding program had no impact on either children's hemoglobin level or anemia. Detailed regression results can be found in Table A9.2–A9.3.
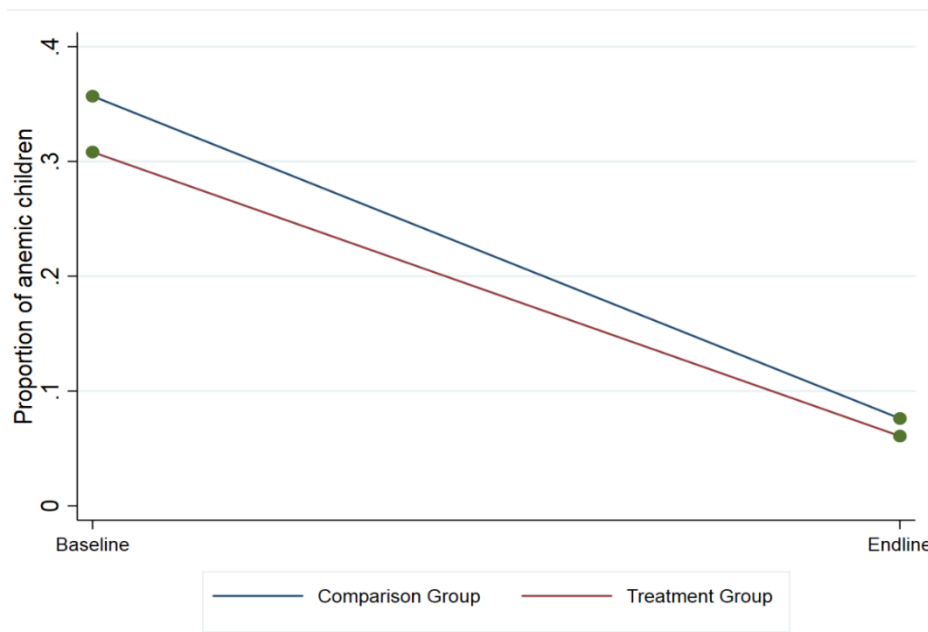
Figure 5 Proportion of anemic children by treatment status and surveys

11.   Although the program had little impact on the Hb level or anemia of average preschoolers, results from subgroup analyses show that the program did help improve the Hb level and thus reduce anemia among certain subgroups. As shown in Figure 6, for girls aged 3 years old in baseline, school feeding intervention reduced anemia by 8 percentage points, whereas no significant effect of the intervention was found in girls aged 5 in the baseline. In contrast, the program reduced anemia by 4 percentage points for boys aged 5 at the baseline survey, but no effect was found in boys aged 3 in the baseline. Detailed regression results can be found in Tables A9.4 in Annex 9.
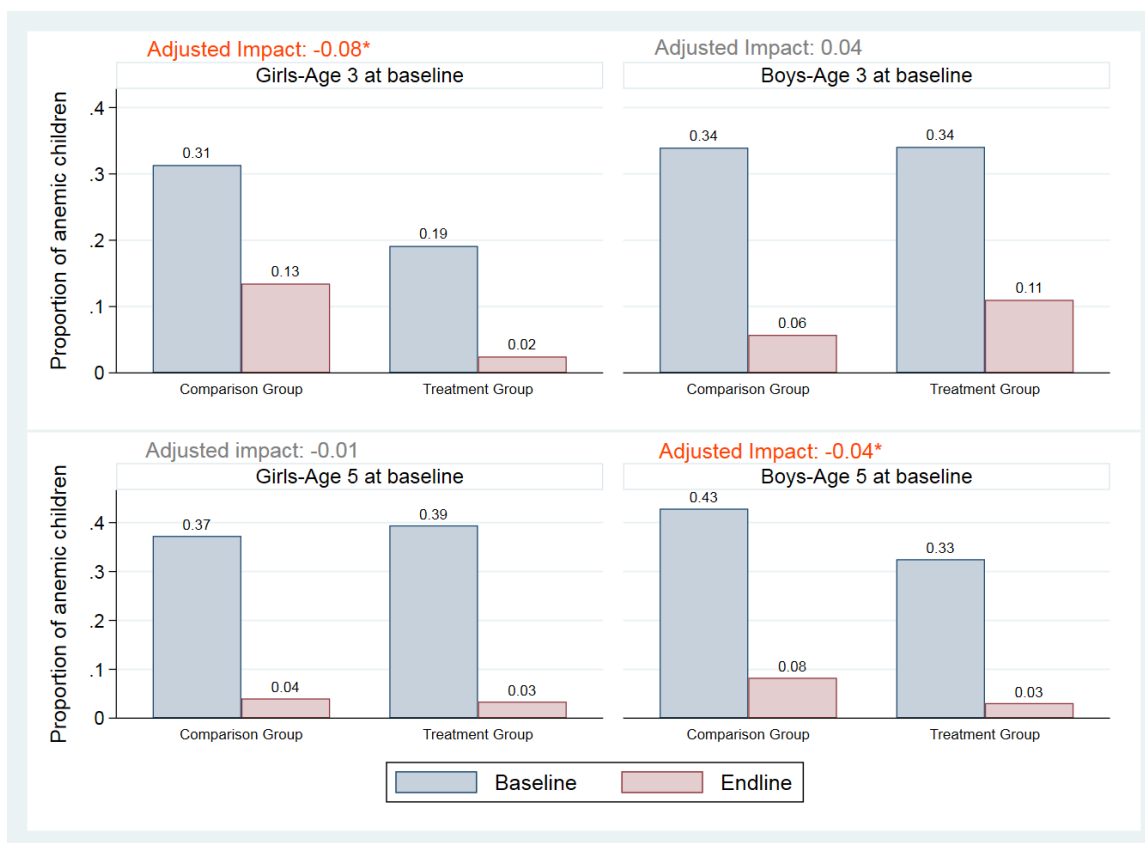
Figure 6 Heterogeneous program impact on the prevalence of anemia, by gender and age

12.　　These results imply that the school feeding program may not be sufficient to reduce the cases of anemia for an average sample preschooler, but they may be effective in reducing anemia for some sub-groups, specifically for those preschoolers who were younger girls or older boys at the baseline survey. However, we did not find any heterogeneous impacts by the SES of preschoolers.

13.　　There are at least three possible explanations for the above findings. First, in line with some literature, impacts on anemia were not detected in randomized controlled trials where the rations did not include multi-fortified foods or iron supplementation (Buttenheim, Alderman, and Friedman 2011; Kazianga, de Walque, and Alderman 2014). Second, similar to other school-based nutrition programs, the program only reached kids who attended school frequently. The most anemic students may not have received enough iron through the meals to build up sufficient iron stores, for example, because of more infrequent school attendance due to the negative effects of anemia such as increased lethargy (Banerjee et al. 2018; Berry et al. 2017). Third, it could also be the case of other diseases like jaundice we do not observe that has a negative effect on the absorption of iron, as well as other minerals and nutrients (Megally and Ghoneim, 2020).

Program impact on children's anthropometric outcomes

14.　　We assessed children's anthropometric outcomes using three standard indices of child anthropometric status – namely, height-for-age Z-score (HAZ), weight-for-age Z-score (WAZ), and weight-for-height Z-score (WHZ). Each of these indices provides different information about growth and body composition, which can be used to assess nutritional status. We constructed most commonly used anthropometric indicators — stunting (low height for age), underweight (low weight for age), wasting (low weight-for-height), and overweight at the other end of the spectrum — based on weight, height, age, and gender with WHO's reference data for "healthy" children. Each of the three anthropometric indices is expressed in standard deviation units (a z-score) from the median of the Multicentre Growth Reference Study sample of children of the same age and sex. The estimated nutritional status of the survey population is expressed as the proportion of children with z-scores below a certain cut-off point. The three anthropometric indices, and corresponding indicators, are further described below.

15.   **Height-for-age Z-score** reflects the linear growth of children. Children who had a HAZ under -2 were considered **stunted**. This index identifies past or chronic malnutrition, which is the effect of long-term poor health, and inadequate diet, which in turn leads to poor linear growth, particularly for children. Figure 7 below shows the distribution of HAZs in the comparison versus treatment groups across two survey rounds. At the endline survey, 4 percent of children in the treatment group were stunted, a decrease from the baseline level of 10 percent. However, the stunting rate decreased similarly in the comparison group, from 12 percent to 6 percent, so that children were still similarly likely to be stunted in both treatment and comparison groups at the endline survey. There is no evidence that the program reduced stunting for an average preschooler in the study area.

16.   Disaggregating the results by sex of children yields different program impacts. In fact, there was an impact of 2 percentage points for boys on reducing stunting prevalence. Detailed regression results can be found in Appendix Tables A11.5- A11.6. As shown in Figure 8 at the endline, 4 percent of boys in the treatment group were stunted, a decrease from the baseline of 12 percent. Boys in comparison group also witnessed a decrease in stunting prevalence, while at a smaller scale (from 11 percent to 6 percent).
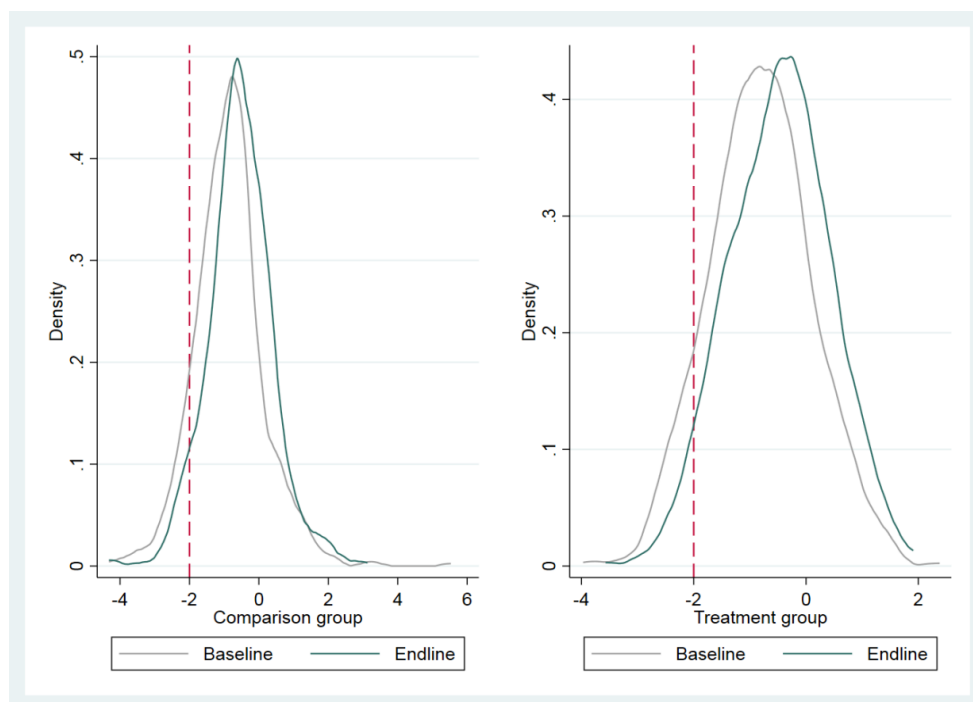
Figure 7 Distribution of height for age z scores
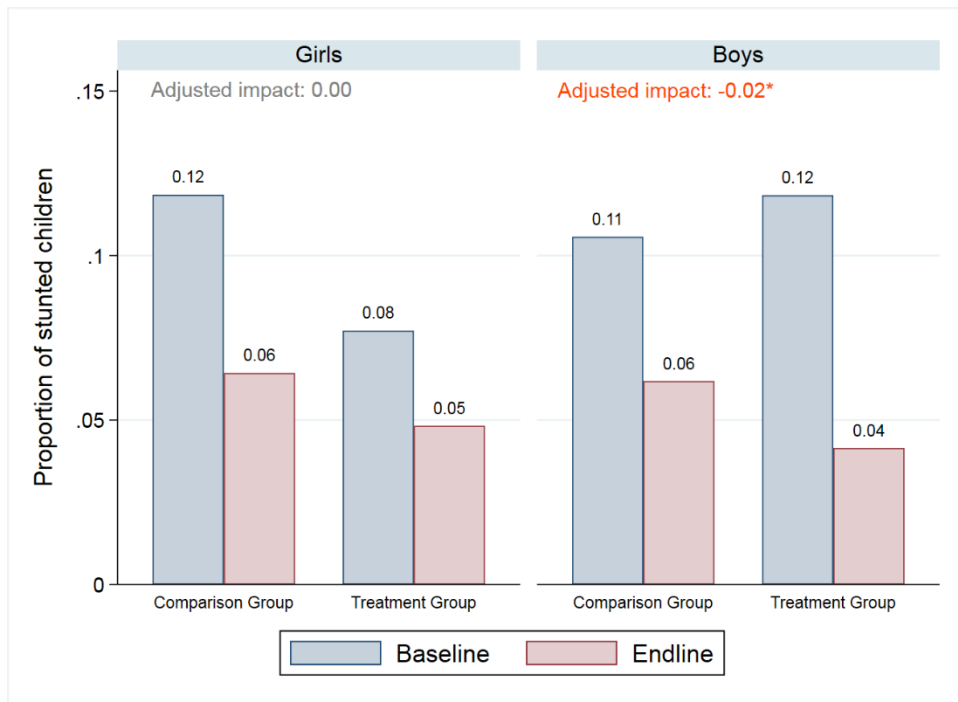
Figure 8 Heterogeneous program impact on the prevalence of stunting, by gender

17.	**Weight-for-age** Z-score reflects body mass relative to chronological age. Children with a WAZ less than -2 are classified as **underweight**, and severely underweight when their WAZ is less than -3. This index reflects both past (chronic) and / or present (acute) undernutrition, although it is unable to distinguish between the two. WAZ analyses indicate a similar story as the previous outcome variables. Like the impacts on stunting, overall, the program also did not reduce the probability of children being underweight (Table A9.5). No significant program impacts were detected in any subgroups, either.

18.	As shown in Figure 9, both groups experienced a decline in the proportion of underweight children from the baseline, where the prevalence is already quite low (6 percent in the treatment group and 5 percent in the comparison group). About 5 percent of treatment children were underweight at the endline, whereas the average underweight rate in the comparison group is lower, at 3 percent.
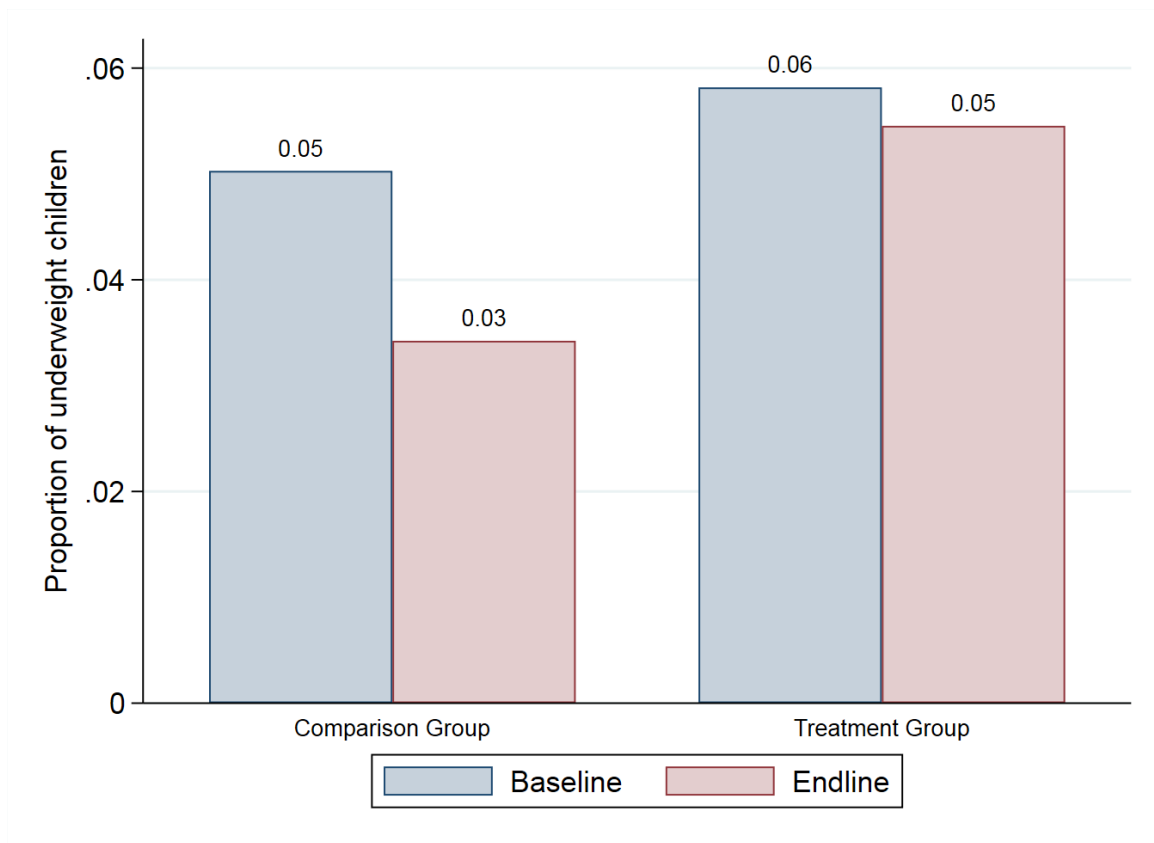
Figure 9 Prevalence of underweight among children

19.   **Weight-for-height** Z-score reflects body weight relative to height. Having a WAZ low less than -2 is referred to as **wasting,** which is attributed to acute malnutrition. The prevalence of wasting at the baseline was 2 percent. As shown in Figure 10, the prevalence increased to 3 percent from the baseline to the endline. This increase in wasting was unexpected considering both stunting and underweight were decreasing. In fact, the project couldn't trace any acute stresses that may affect weight but not height. However, we refrain from concluding too much from this increase given that the increase was only minimal.
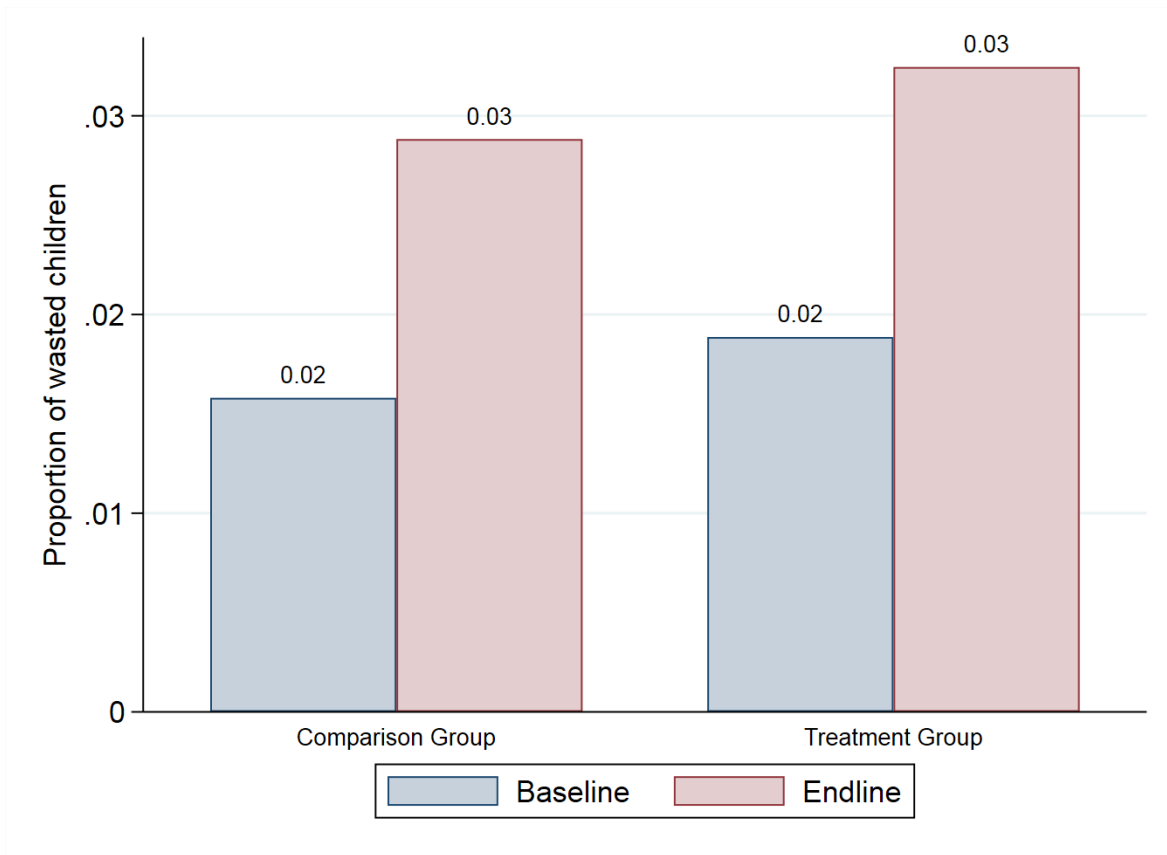
Figure 10 Proportion of wasted children, by treatment status

20.   Though we could not find any effect on wasting of an average sample preschooler, results from sub-group analyses show that the program led to a decline of 4 percentage points in the prevalence of wasting among children from a registered poor household (Table A9.7). As shown in Figure 11, the effect is mainly driven by a larger increase in wasting of the comparison group from the baseline to the endline.

21.   At the other end of the spectrum, it appears that the program does have a nontrivial impact on the likelihood of being overweight by 6 percentage points, especially for children from poor households. As shown in Figure 11, for children from poor households, the prevalence of overweight in the treatment group increased from 11 percent at the baseline to 18 percent at the endline. These results are consistent with the literature that found a positive association between school meals and overweight, although they were mostly conducted in high income countries. For example, Schanzenbach (2009) found that children who consume school lunches are more likely to be obese than those who bring their lunches even though they enter preschool with the same obesity rates. These results are also consistent with previous findings that School Breakfast Program (SBP) and National School Lunch Program (NSLP) exacerbates childhood overweight and obesity in the United States (Millimet et al., 2010; Capogrossi and You, 2017).
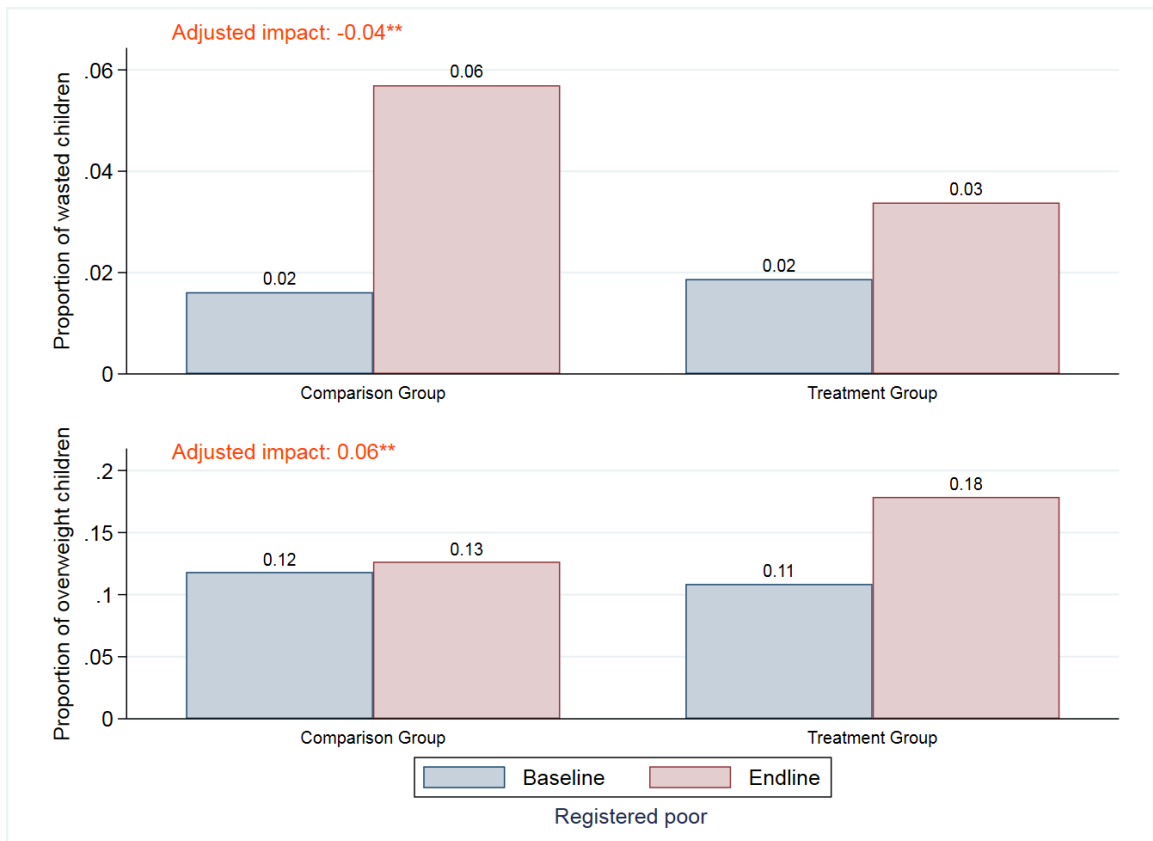
Figure 11 Program impact on the prevalence of wasting and overweight for registered poor

22.   Overall, we found no impact on the anthropometric measures or anemia of average preschoolers in the study site. The evaluation did find some evidence of impacts of the intervention on children's nutritional status of some sub-groups in terms of child's age, gender, and poverty status. While results have not been able to provide unified support for school feeding impact, it has pointed out the need to look at underlying mechanisms and implementation challenges. For example, there might be some supply-side bottlenecks holding back the intervention from realizing its full potential. The nutritional value of the food choices offered in schools may have an impact on the potential of improving some children's nutrition and the chance that they become overweight, thus needs to be further examined.

23.   Note that the Project was evaluated against a backdrop of China's national momentum towards reducing child malnutrition and widespread improvements in key maternal and child health indicators. This was also one of the explanations that the project did not show impacts as child nutrition in both treatment and comparison groups are improving, possibly thanks to the national efforts. However, even in the context of these achievements, there is still a great deal of room for improvement in many indicators, as evidenced by our comparison between sample children and national averages on indicators such as stunting, underweight, and additional attention is needed on the observed rise in the prevalence of overweight and obesity.

2.1.2. SUB-QUESTION 2: TO WHAT EXTENT DOES THE NUTRITION PROGRAM IMPROVE PRESCHOOLERS' COGNITIVE ABILITY?

Introduction of the test instrument

24.　　Cognitive ability was assessed using a battery of two sub-tests taken from the Mandarin-language version of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-IV) and the Wechsler Intelligence Scale for Children Fourth Edition (WISC-IV). The WPPSI-IV was used to measure the cognition of sample children aged from 4 years old to 6 years and 11 months whereas the WISC-IV was used to measure the cognition of sample children aged 7 years and above. As the latest versions, both WPPSI-IV and WISC-IV were culturally adapted, translated and edited into simplified Chinese and validated for Chinese children based in 2008. Since research suggests that children's verbal comprehension and working memory are those areas of cognitive ability most likely to be affected by nutrition, we focused our efforts on measuring these two outcomes.

25.　　In WPPSI-IV, the Verbal Comprehension Index (VCI) is assessed through two core subtests: Information and Similarities subtests. The working memory index (WMI) is assessed through two other core subtests: Picture Memory and Zoo Location. Similarly, in WISC-IV, the VCI is assessed through two core subtests: Coding, and Symbol Search. The WMI is assessed through two other core subtests: Digit Span, and Letter Number Sequencing. Raw scores obtained from these core subtests were converted to age-scaled index scores using tables of norms from the official WPPSI-IV and WISC-IV administration and scoring manuals for China, respectively. According to the scoring manual, index scores 70-130+ indicated 7 performance levels from extremely low to very superior, with scores from 90 to 109 considered average, 80 to 89 considered Low average, 70 to 79 considered Borderline, less than 70 considered Extremely Low Average. Children whose index scores fall in the 70-79 (Borderline) range performed worse than approximately 91% of children the same age and gender in the WPPSI-IV and WISC-IV normative sample and might have some learning disabilities. Children whose index scores fall in the below 70 (Extremely Low) range performed worse than about 98% of children the same age and gender and might need to be placed in special education courses. Sample children was individually administered the four core sub-tests by trained examiners.

Children' s Performance in Cognitive Test in the Endline

Table 3 Disaggregated data for child cognitive ability at the endline

	Average VCI	Low VCI [80-89] (%)	Borderline VCI [70-79] (%)	Extremely Low VCI [<70] (%)	WMI	Low WMI [80-90] (%)	Borderline WMI [70-79] (%)	Extremely Low WMI [<70] (%)
overall	89.21	25.37	21.26	4.02	91.66	22.55	21.06	1.30
Age								
4-6 years old	90.43	30.79	14.32	0.72	99.45	15.27	5.73	0.48
7-years and older	88.32	21.43	26.24	6.45	86.06	27.79	32.08	1.89
County								
Longshan	90.72	23.09	19.83	2.33	91.48	22.76	20.46	1.46
Yongshun	87.85	27.45	22.56	5.57	91.82	22.37	21.61	1.15
Treatment Status								
Treatment	88.64	24.44%	22.70	4.93	91.75	23.11	21.06	1.43
Comparison	89.77	26.28%	19.88	3.16	91.58	22.03	21.05	1.17
Gender								
Female	89.05	25.87	21.96	4.31	91.09	20.93	21.95	1.42
Male	89.37	24.90	20.53	3.75	92.21	24.12	20.20	1.18
Ethnicity								
Han	91.52	13.59	23.58	3.88	90.90	25.47	18.87	2.83
Non-Han	88.95	26.74	20.98	4.04	91.75	22.20	21.32	1.11
Household Poverty Status								

	Average VCI	Low VCI [80-89] (%)	Borderline VCI [70-79] (%)	Extremely Low VCI [<70] (%)	WMI	Low WMI [80-90] (%)	Borderline WMI [70-79] (%)	Extremely Low WMI [<70] (%)
Non-Poor	90.00	24.59	21.13	2.93	92.03	22.42	19.84	1.13
Poor	87.95	26.65	21.47	5.80	91.05	22.77	23.04	1.57
Among Poverty Status: Registered Poor								
Non-Poor	90.06	25.16	20.59	2.81	92.07	22.45	19.81	1.08
Poor	87.68	25.78	22.47	6.23	90.91	22.75	23.32	1.69
Left-Behind Status								
Non-Left-Behind	88.92	22.71	21.81	5.08	92.33	20.47	21.81	0.34
Left-Behind	89.34	26.50	21.02	3.58	91.38	23.44	20.74	1.70
Father's Education								
Below Senior-High	88.30	25.73	23.33	4.60	90.53	24.79	23.13	1.88
Above & Senior-High	90.06	25.05	19.35	3.50	92.70	20.50	19.16	0.77
Mother's Education								
Below Senior-High	87.57	28.60	24.27	5.18	90.02	24.72	25.84	1.35
Above & Senior-High	90.54	22.77	18.85	3.10	92.97	20.83	17.24	1.26

26.   Data from the endline survey (Table 3) showed that the VCI and WMI scores of sample children averaged 89.21 (Low Average level) and 91.66 (Average level), respectively. Within 50.65% children, who had a VCI lower than average level (<90), 4.02% of them fell in the range of extremely low level (<70). Comparing to the VCI, WMI scores look better, as 44.91% children had WMI lower than average level (<90), and 1.03% in the range of extremely low (<70). Children in the lower VCI ranges might face difficulties when handling language-type tasks, at risk for listening comprehension, verbal reasoning, and oral responding. Lower WMI scores imply that children had problem in concentrating and attending to information that is presented to them. For this type of children, processing of complex information might be more time consuming compared to their peers, potentially resulting in difficulties in a variety of learning tasks.

27.   Results from sub-group analyses show significant differences between different age groups. Specifically, both the **WMI and VCI of 4-6 years old children were significantly higher** than those of children aged 7 and older. The proportion of low VCI among age 4-6 was 30.79%, significantly higher than that of age 7 and above (21.43%). In contrast, the proportion of low WMI of the former was significantly lower than that of the latter (15.27% versus 27.79%). Children whose parents' education was lower than senior high school **scored significantly lower in both VCI and WMI** than their counterparts whose parents graduated from senior high school or above. Similarly, data also show that VCI varies by countries, ethnicity and SES of the children. Specifically, non-Han children who came from low SES households in Yongshun county scored much lower VCI than otherwise. However, no significant differences in VCI was observed by children's treatment status or gender.

28.   We began with descriptive analyses by comparing children's cognitive ability between comparison and treatment groups from baseline to endline (Figure 12). Our data show that children's VCI and WMI score have increased 3.81 and 1.56 respectively in the comparison group and have increased 2.57 and 0.88 respectively in the treatment group from baseline to endline. Simple mean difference test shows that the gain in VCI from baseline to endline is significantly higher in the comparison group than in the treatment group. However, the gain in WMI is not statistically different between the two groups.
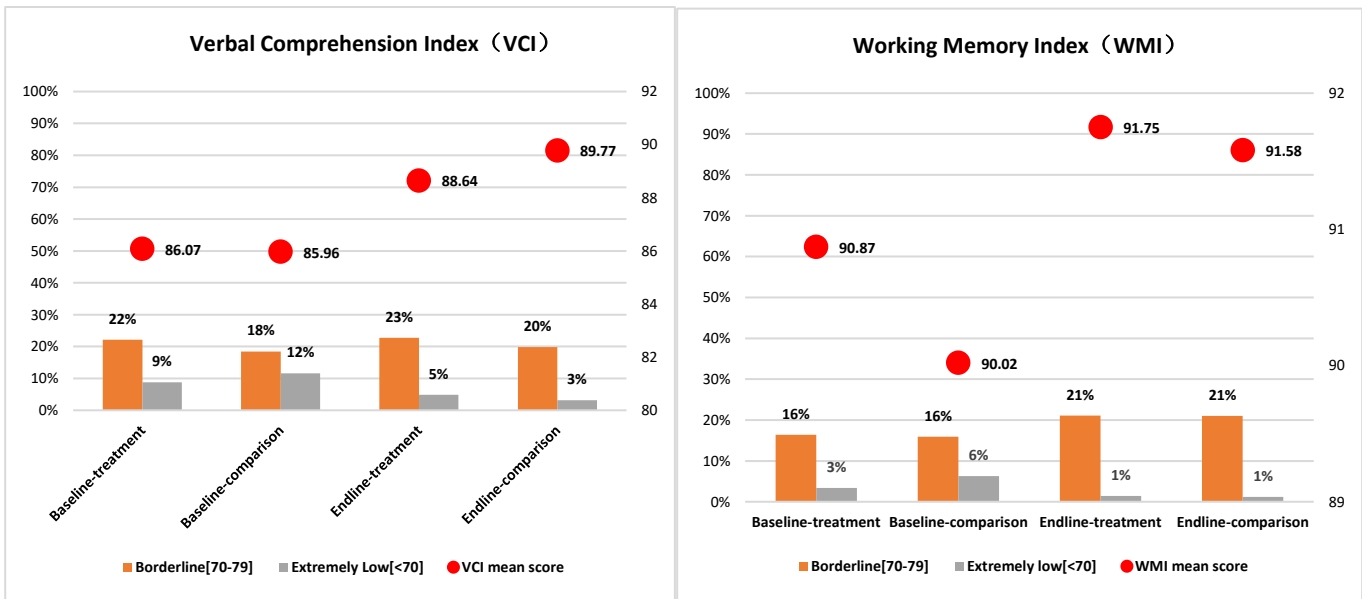
Figure 12 Changes on children's cognitive ability by treatment status from Baseline to Endline

Program impact on social-emotional status

29.   We then conducted multivariate analyses to further estimate the impact of the program on children's cognitive ability. To do so, we took a two-step approach. First, we estimated the Intention to Treat (ITT), which means we estimated the program impact based on the initially assigned treatment status of children and did not take into account their actual treatment status. Secondly, considering the existence of cases of non-compliance, we also estimated Treatment on the Treated (TOT) based on the actual treatment status of children. Cases of non-compliance include some children who were initially assigned into the treatment group ended up not receiving treatment (say because the preschools they attended in the baseline survey was closed and they got transferred to a non-treatment preschool), whereas some children who were initially assigned into the comparison group ended up receiving treatment (say because one comparison preschool was switched into treatment preschool by the local project office after the baseline). As a robust check, we run three different empirical models: treatment dummy variable only, treatment dummy variable plus the baseline level of outcome variable, treatment dummy variable plus baseline level of outcome variable as well as other control variables. As can be seen from Table A9.8 and A9.11 in Annex 9, results from both ITT and TOT analyses show that the preschool nutrition pilot has little effect on children's cognitive ability in terms of VCI and WMI.

30.   Similarly, results from ITT and TOT analyses also show that the program has little impact on to the incidence of Low and Extremely Low VCI/WMI scores either (Table A9.9, Table A9.10, Table A9.12 and Table A9.13 in Annex 9).

Disaggregated data by gender, age, ethnicity, school level, poverty status, left-behind status and educational level of parents

31.   In order to explore whether the program has impact on certain groups of children in terms of cognition, we conducted sub-group comparisons. As is shown in Figure 13, among the sample children from poor families, **the comparison group experienced more gain in VCI from baseline to endline than the treatment group**. Why is it like this? According to Hair et al. (2015), the progress of cognitive ability was related to the structure of brain development which has close relationship with family poverty status. Some factors such as parenting styles, grow up stresses and mental health are more important to the development of children's cognitive abilities rather than nutrition (Guo & Harris, 2000). Children in the poorest environment has lower cognitive ability especially language use ability than their peers from better-off environment (Perkins et al., 2013).
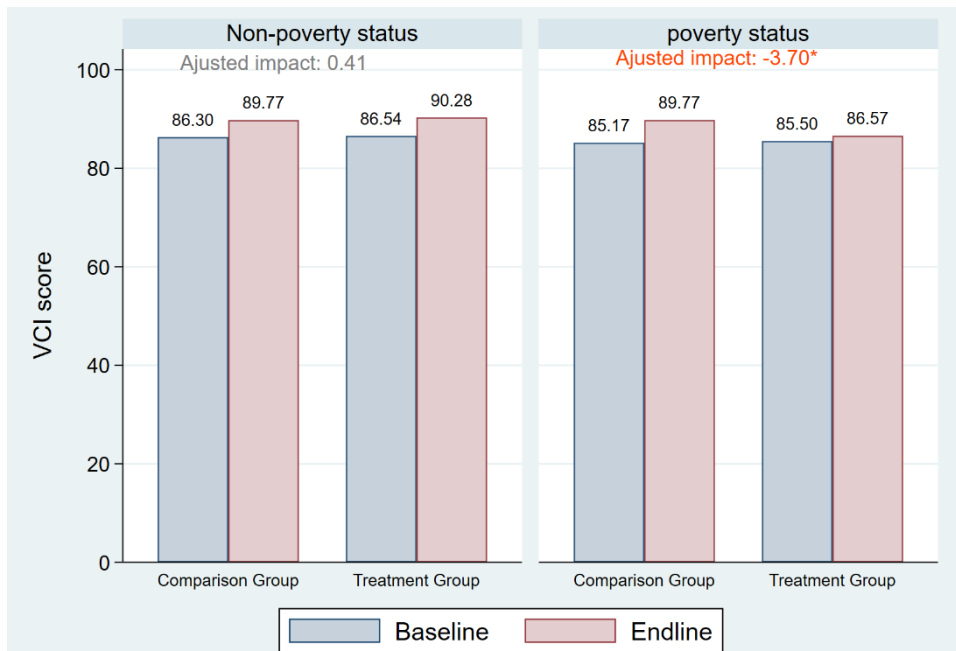
Figure 13 The program's impact on children's VCI score in different poverty status

32.　　　Furthermore, we found the consistent results when dividing the samples into five different groups according to the wealth index (Figure 14). The nutrition program has negative impact on the VCI score of the child who grows up in the poorest conditions (coefficient is -3.25, p<0.1) and the proportion of Low WMI children significantly increased (p<0.05), while the proportion of Low WMI and VCI children significantly decreased when the child was in the environment of richer families (Low WMI: p<0.1; Low VCI: p<0.05). The other heterogeneous analysis results can be found in Table A9.15 in Annex 9.
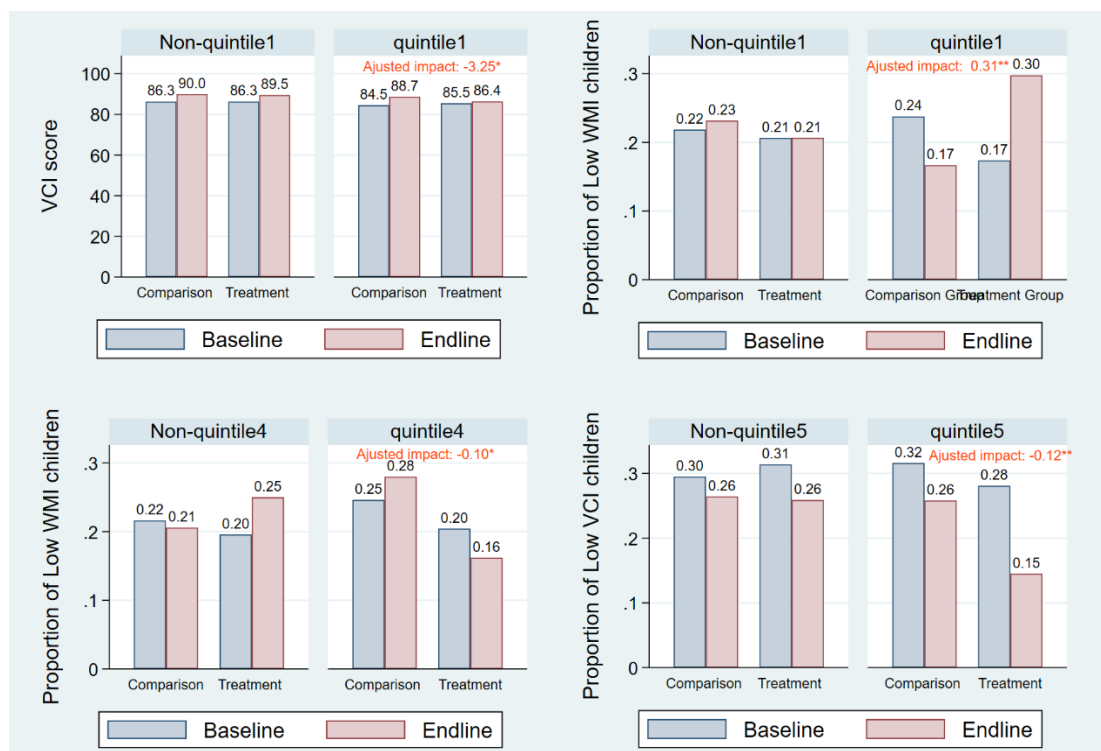
Figure 14 Heterogeneous analysis of the impact on children's cognitive ability in different groups

Conclusion

33.   Endline data showed that the VCI and WMI scores of sample children averaged 89.21 (Low average level)) and 91.66 (Average level) respectively. Comparing with the norm proportion of Extremely Low (2.2%), a total of 4.02% and 1.03% children had a VCI and WMI that was Extremely Low, an underperforming result compared with average children at the same age.

34.   With the VCI and WMI mean score itself has increased across the time flying from baseline to endline, the proportions of abnormal cognitive ability have decreased a lot.

35.   The significant differences were detected between children's different age groups (p<0.01), county groups(p<0.05) and the children whose parents' education level below and above the senior high schools (p<0.01). And the data shows children from poorer family scored lower than the one from richer.

36.   The ITT and TOT analysis uncovered that the nutrition program has no significant influences on the children's cognitive ability. The heterogeneous analysis showed that the children from the poorest family may have worse cognitive ability related to the nearly stagnant brain structure development. The children from better family wealth condition may have benefited from the nutrition project as the proportion of Low VCI and Low WMI children decreased significantly with the nutrition improved.

2.1.3. SUB-QUESTION 3: TO WHAT EXTENT DOES THE NUTRITION PROGRAM IMPROVE PRESCHOOLERS' SOCIO-EMOTIONAL STATUS?

37.   Preschool education is the key period of children's social-emotional development. One of the objectives of this part is to examine the impact of the school feeding program on children's social-emotional status. We first introduced the two main testing instruments, the Strengths and Difficulties Questionnaire (SDQ) and the Age & Stage Questionnaire: Social-Emotional (ASQ:SE), and then further described the abnormal rates of children's socio-emotional status. We further showed the effects of the program on overall children and subgroups of children. Finally, we discussed the program impacts.

Introduction of the test instrument

38.   In the endline survey, two scales were administered by trained enumerators face to face to caregivers as part of the caregiver questionnaire to measure sample children's social-emotional development, including mental health status, as well as their behaviors and emotions.

• The Strengths and Difficulties Questionnaire (SDQ)

39.   The Strengths and Difficulties Questionnaire (SDQ) was applied to capture sampling children's mental health status in the endline survey. As a widely accepted measurement tool for child psychological development, SDQ has been translated into different languages including Chinese. It contains 25 questions which test different dimensions of child's social emotions, including emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems and prosocial behaviors. Negative attributes were laid out for assessing the former four aspects while positive behavior items were used to measure the last division of children's social emotions.

40.   **Implementation Process:** In the endline survey, trained enumerators conducted face to face interviews with caregivers of the sample children and completed these 25 questions in SDQ. Positive attributes like "Considerate of other people's feelings" as well as negative attributes like "Restless, overactive, cannot stay still for long" were laid out in the questionnaire. The respondent would choose from three options to best describe the child's performance, including "Not true", "somewhat true", and "certainly true".

41.   **Scoring:** Five subset scores were calculated based responses to caregivers to show children's emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems and prosocial behaviors respectively. The first four scores (emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems) were regarded as measurements of difficulties. By aggregating them, the team calculated the total difficulties score for each sample child. Except for prosocial behaviors, higher score always means worse social-emotional status.

42.   **Borderline:** By comparing those raw scores with corresponding borderlines, the team could identify whether this child required further monitoring in terms of psychological health or not. Children scored higher than 4 for emotional symptoms, 3 for conduct problems, 7 for hyperactivity/inattention, 5 for peer relationship problems or 16 for total difficulties (sum of above four), respectively, might be lagged behind

compared to their peers in these aspects. Moreover, sub-score for prosocial behavior lower than 5 reported abnormal.

- Age & Stage Questionnaire: Social-Emotional (ASQ: SE)

43.   The endline survey also administrated Age & Stage Questionnaire: Social-Emotional (ASQ: SE) to screen children's emotions and behaviors. ASQ: SE, which consists of different age-specific questionnaires, has been widely used in the literature to identify the likelihood of children anxiety, depression and even some anti-social behaviors. This scale has gained more and more popularity in fieldwork because of its user-friendliness and cost-effectiveness.

44.   **Implementation Process:** Because our sample children are all older than 5 years old, we used the 60 months/5 years Questionnaires for all the sample children. The child's caregiver was asked to answer the multiple-choice questions from the selected questionnaire one by one without any disturbance. For example, items like "Does your child look at you when you talk to him" were set and the caregiver would choose the best answer from "most of the time" "sometimes" or "rarely or never" according to the child's everyday behavior and performance. We also asked whether this item would be a concern to the interviewee and got a Yes/No answer.

45.   **Scoring:** When scoring the questionnaire, each choice would respectively refer to 0, 5, or 10 points, respectively. And if the item became a concern, there would be an extra 5 points. By adding up scores of all the questions, we got a total raw score which would be compared with cutoff values later to demonstrate whether this child need further screening. In other words, whether a child has any potential social-emotional problems.

46.   **Cutoff values:** Since the team only used one set of scales, the cutoff points were the same for all the children. In reference to the User's Guidebook, the team set cutoff values at 70 for 60months/5years Questionnaire.

Percentage of Children Scored above the Cutoff Value

47.   In terms of six subset scores of SDQ, Figure 15 depicted that 10% of the children might need further concern in total difficulties. To be specific, for peer relationship, only 3% of the sample exceeded the borderline. Children underperformed in conduct problems and hyperactivity, both accounted for 9%. In addition, 15% of the children were abnormal in emotional symptoms. Besides, 6% of the children were less likely to exert prosocial behaviors. On the contrary, another study showed that among a sample of 3 to 17-year-old Chinese children, 12.1% were screened out to be abnormal according to total difficulties score and 9.1% to have some difficulties in prosocial behaviors.

48.   Figure 15 also depicted that 41% of the children got a higher ASQ:SE score when compared to the corresponding cutoff value, which reflected that nearly 2/5 of the children might have potential social-emotional problems. In contrast, a study conducted a Chinese city, Xiamen, only reported a 23.7% probability to be emotional delayed among 239 children aged 38 to 66 months.
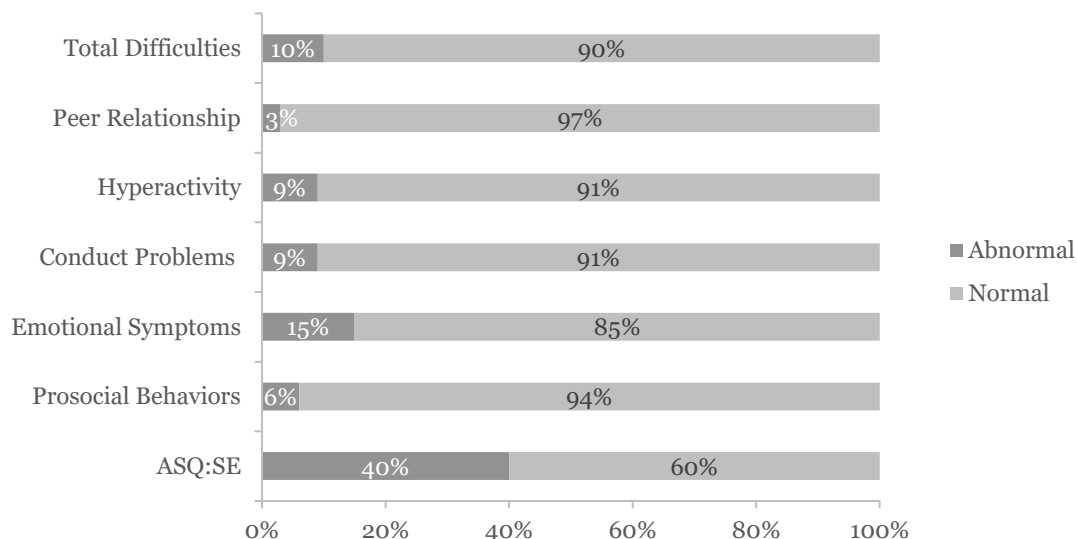
Figure 15 Percentage of children scored above the cutoff value in socio-emotional test

Disaggregated data by gender, age, ethnicity, school level, poverty status, left-behind status and educational level of parents

49.   In this part, we reported socio-emotional abnormal rate of different subsamples of endline data to see if there were inherent disparities existing between different groups, which might point out some significant factors related to child social-emotional status. The results were shown in table1.

50.   Endline data (Table 4) illustrated that children in both counties performed similarly in most dimensions of social emotion status. Specifically, the abnormal rate for prosocial behaviors and the deviant rate for total difficulties among sample children were statistically the same in the two counties, However, the abnormal rate of ASQ:SE of children was significantly lower in Longshan than that in Yongshun county (36% vs 46%).

51.   Table 4 also illustrated that children in treatment group performed as well as those in comparison group in social emotion status since their means are statistically the same.

52.   **Moreover, disaggregated data showed disparities in gender or age but not in ethnicity or education levels.** Specifically, boys performed significantly better in total difficulties than girls. When it comes to prosocial behaviors and ASQ:SE score, the gender gap was less obvious. Difference between age groups was similar to that between gender groups. However, data disaggregated by ethnicity could not provide any strong evidence that ethnicity difference existed. The difference between children from primary schools and preschools was also insignificant.

53.   Finally, we shed some light on the potential relation between family background and children's socio-emotional status. Children from poor family were more likely to be abnormal in prosocial behaviors. However, children from registered poor families performed better in ASQ:SE score than those from otherwise families. However, children with different household economic backgrounds did not differ in other dimensions of socio-emotions. Besides, our data did reveal significant difference in social-emotional development between left-behind and no left behind children. However, in terms of education level of parents, children with better educated mothers tend to perform better in total difficulties.

Table 4 Abnormal rate of child social emotional status at endline

	SDQ		ASQ:SE
	Total Difficulties	Prosocial Behaviors	Raw score
Overall	10%	6%	40%
County			
Longshan	9%	6%	36%
Yongshun	11%	5%	45%
Treatment Status			
Treatment	9%	6%	39%
Comparison	10%	5%	41%
Gender			
Male	12%	5%	42%
Female	8%	6%	38%
Age			
7-years and older	8%	6%	40%
4-6 years old	11%	6%	40%
Ethnicity			
Non-Han	10%	6%	40%
Han	11%	8%	39%
School level			
Primary school	9%	5%	41%
Preschool	11%	6%	39%
Poverty status			
Poor	11%	6%	40%
Non-Poor	9%	6%	40%
Among Poverty Status: Registered poor			
Poor	11%	6%	41%
Non-Poor	9%	6%	40%
Left-behind status			
Left behind	10%	6%	41%
No left behind	9%	4%	38%
Mother's educational level			
Below Senior-High	11%	6%	41%
Above & Senior-High	9%	5%	40%
Father's educational level			
Below Senior-High	12%	6%	41%
Above & Senior-High	8%	6%	41%

Program impact on social-emotional status

54.   Figure 16 showed the distribution of three indicators of children's social-emotions in endline survey, and the red line marked the cut-off point. In terms of total difficulty score, the probability distribution of the treatment group and the comparison group was similar, but the score of the treatment group was slightly lower than that of the comparison group. Similarly, ASQ:SE showed a similar distribution between the treatment and comparison groups. However, in terms of prosocial behavior, the comparison group scored much lower than the treatment group.
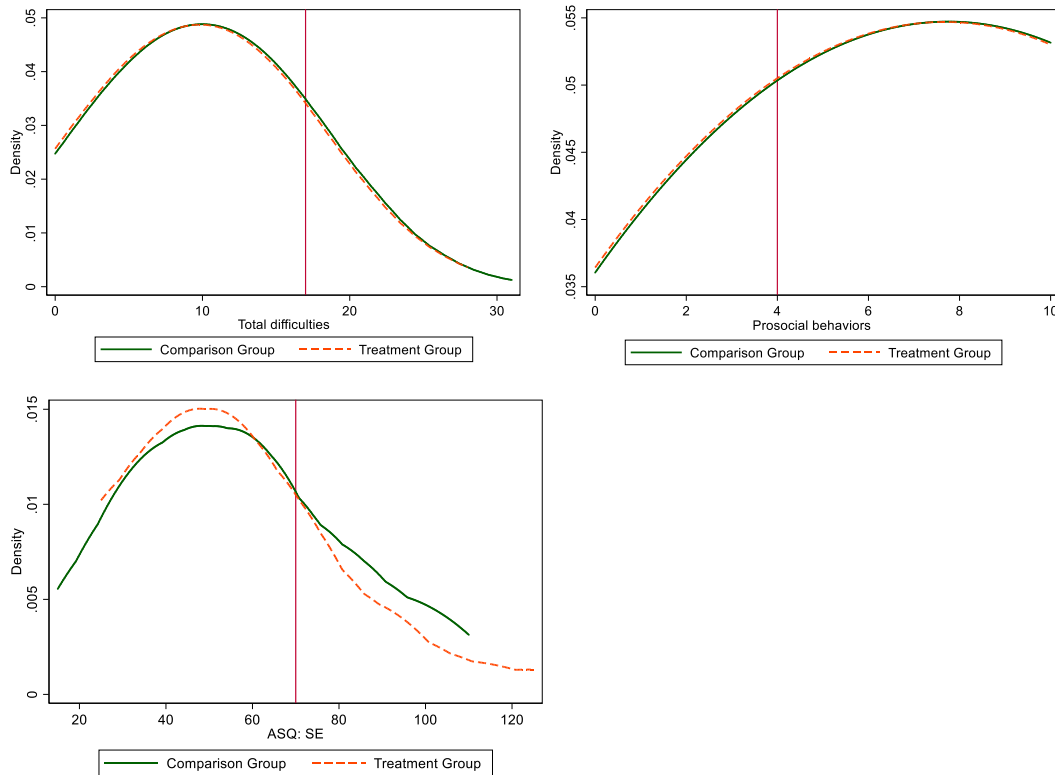
Figure 16 Distribution of children's social-emotional status in endline by treatment status

55.   In terms of the abnormal rate, from baseline to endline, children in both the treatment and comparison groups showed a large decrease in all measures of social-emotional status. To be specific, children in the treatment group declined more on the abnormal rate of total difficulties and ASQ:SE, while the comparison group saw a larger decline in prosocial behaviors. Figure 17 below shows how this proportion changed in the two groups. In total difficulties, around 20% of sample children showed abnormal in the baseline survey. The treatment group saw a large decline from 21% in baseline to 9% in endline, and the comparison group showed a smaller decline from 19.3% to 10.5%. A similar reduction is shown in prosocial behaviors, while the abnormal rate of comparison group (from 15.9% to 5.8%) declined a little faster than that of treatment group (from 15.0% to 6.3%). In ASQ:SE, around 70% of sample children showed abnormal in the baseline survey, and the treatment group saw a larger decline from baseline to endline survey, from 70% to 40%, compared to the downward trend of comparison group (from 69% to 41.6%).
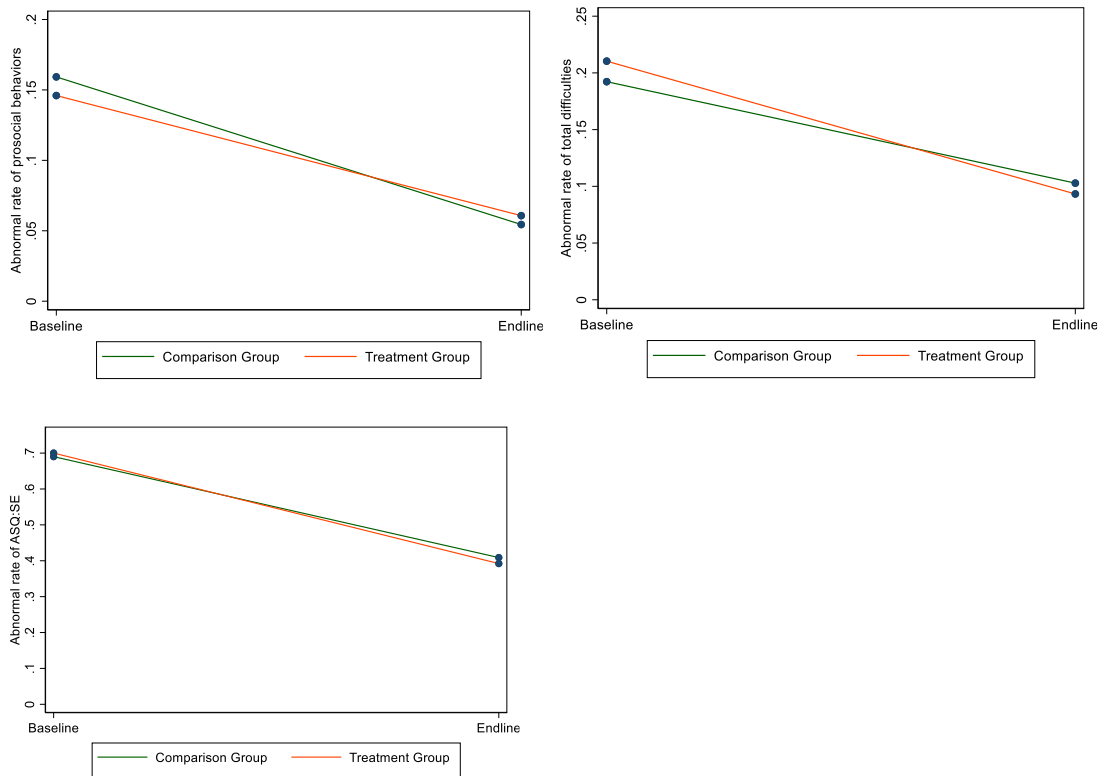
Figure 17 Abnormal rate of children's social-emotional status across baseline and endline, by treatment status

56.   The program impact on children's social emotions was shown in Tables A9.16-A9.21 in Annex 9. These results implied that the school feeding program could reduce the total difficulties (as measured by both level and abnormal rate) and the abnormal rate of social emotional status measured by ASQ:SE. And it also had some impact on the abnormal rate of hyperactivity in SDQ (significant at 5%).
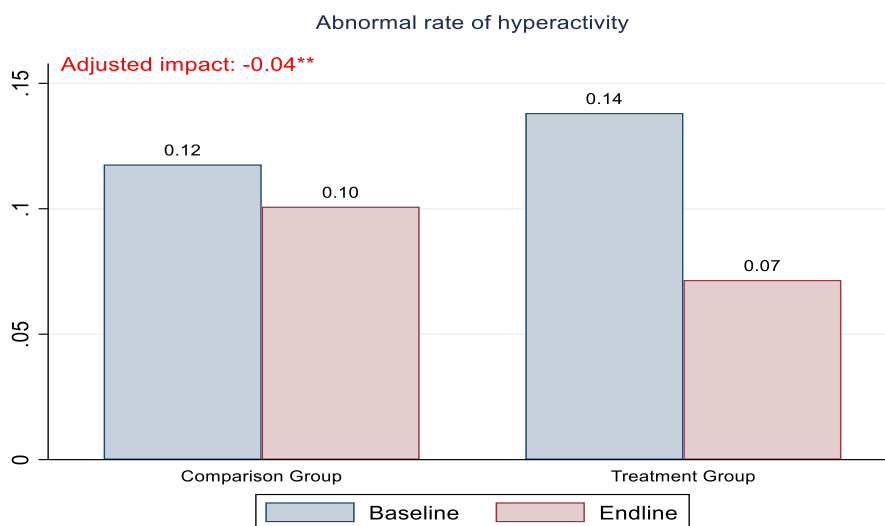
Figure 18 Program impact on the abnormal rate of hyperactivity

57.   Although the school feeding program did not have a significant effect on most social emotional indicators, it may improve the status of some subgroups on some indicators of the total difficulties (shown in Table 5). The detailed regression results were shown in Tables A9.22.

58.   Results from regression analyses by subgroups show that for boys, school feeding intervention reduced the abnormal rate of emotional symptoms by 0.06 percentage point, whereas no significant effect of the intervention was found among girls. In contrast, the school feeding program reduced the abnormal rate of hyperactivity for girls by 0.03 percentage points, but no impact on that of boys.

59.   Besides, the younger children aged 3 years old in the baseline could benefit more from the school feeding intervention. The intervention reduced the abnormal rate of hyperactivity of the younger children and thus reduce their abnormal rate in total difficulties. However, for those children aged 5 years old in baseline, the intervention increased slightly their probability of showing abnormal in peer relationships.

60.   Children from then poor households also benefited more from the intervention. Similar to the baseline report, we define children's families as poor if they meet the following conditions: 1) registered poor households; 2) subsidized families; 3) low-income families (their income is under the poverty line). For these children, school feeding program improved their emotional symptoms and hyperactivity problems significantly.

Table 5 Program impact on the abnormal rate of total difficulties by gender, age and poverty status

	Coefficient or Odds Ratio (95% CI)	P-value	Coefficient or Odds Ratio (95% CI)	P-value
	Younger Children (*N=501*)		Elder Children (*N=456*)	
Total Difficulties	-0.07** (-0.11, -0.01)	0.012	0.03 (-0.11, 0.08)	0.139
Emotional Symptoms	-0.05 (-0.12, 0.01)	0.109	-0.01 (-0.08, 0.07)	0.866
Conduct Problems	-0.03 (-0.08, 0.02)	0.215	0.02 (-0.02, 0.06)	0.356
Hyperactivity	-0.06** (-0.11, -0.01)	0.024	-0.01 (-0.05, 0.02)	0.431
Peer Relationship	-0.01 (-0.04, 0.02)	0.477	0.03** (0.00, 0.06)	0.040
	Poverty households (*N=358*)		Non-Poverty households (*N=599*)	
Total Difficulties	-0.03 (-0.09, 0.03)	0.381	-0.01 (-0.06, 0.03)	0.595
Emotional Symptoms	-0.07** (-0.15, 0.00)	0.053	-0.01 (-0.07, 0.05)	0.782
Conduct Problems	0.01 (-0.06, 0.07)	0.865	-0.01 (-0.05, 0.03)	0.684
Hyperactivity	-0.05* (-0.11, 0.01)	0.088	-0.03 (-0.07, 0.11)	0.143
Peer Relationship	0.03* (-0.01, 0.07)	0.098	0.00 (-0.02, 0.03)	0.820

Conclusion

61.   In general, these results imply that although the school feeding program may not be sufficient to improve the prosocial behaviors and ASQ:SE of the sample children, but it could help reduce their abnormal rates of hyperactivity. Besides, the intervention may be effective in reducing the abnormal rate of total difficulties for some sub-groups. Firstly, the school feeding program improved hyperactivity of girls and emotional symptoms of boys. Secondly, the younger children (aged 3 in the baseline survey) benefited more from the intervention, especially in hyperactivity. Finally, children from poverty households benefited more from the school feeding program, these effects were mainly reflected in emotional symptoms and hyperactivity.

62.   There are at least three possible related explanations. First, in line with some literature, the impact of nutrition interventions or education on children's mental health was generally not significant (Ajie & Chapman-Novakofski, 2014; Larson & Yousafzai, 2017), although two studies found that among adults, interventions that promoted healthy eating could promote recovery with mental health problems (Cherak et al., 2020; Whybird et al., 2020). Second, studies in the field of child development confirmed that there was a significant correlation between preschool quality and children's socio-emotional development (Burchinal et al., 2010; Hestenes et al., 2015), our future analyses should pay more attention to preschool quality in more depth. Finally, this program only involved those children who attended preschools, and those who did not attend preschools may be at a greater nutritional disadvantage, so not taking these children into account may have skewed our results downward.

63.   In next section, we explore the potential underlying mechanisms behind the lack of impact of this school feeding impact on children's nutrition, cognition and social emotions. To do so, we follow the Theory of Change mapped out in Annex 3 and focus on three critical program outputs: feeding practice, nutrition knowledge and interactive parenting. The evaluation sub-questions were as follows:

2.1.4. SUB-QUESTION 4: TO WHAT EXTENT DID THE NUTRITION PROGRAM IMPROVE THE NUTRITION KNOWLEDGE OF PRESCHOOL PERSONNEL AND CAREGIVERS?

Nutrition knowledge of caregivers

64.   To maintain the panel nature of the data, caregivers were asked the same 11 questions on nutrition and young child feeding knowledge in the endline survey as in the baseline survey. A total of 915 caregivers completed the module, of which 487 and 428 in the treatment and comparison groups, respectively. We then aggregated these responses into nutrition knowledge scores. One point was given for each correct response. If there was more than one correct choice, the respondent could get one point after answering all the correct choices. We rescale the raw scores to a score ranging between 0 and 100.

65.   Table 6 presented the primary caregivers' nutrition knowledge scores of both treatment and comparison groups. The scores of both groups increased slightly from baseline to endline. Figure 19 below shows the distribution of nutrition knowledge scores of caregivers in the comparison versus treatment group from the baseline to endline. There is no evidence that the program increased caregivers' nutrition knowledge scores for the full sample.

Table 6 Primary caregivers' nutrition knowledge scores

Group	Treatment group (1)	Comparison group (2)	P-value [H0: (1)=(2)] (3)
Baseline	28.1	29.2	0.23
Endline	37.1	37.0	0.89
Change from Baseline to Endline	9.0	7.8	0.31
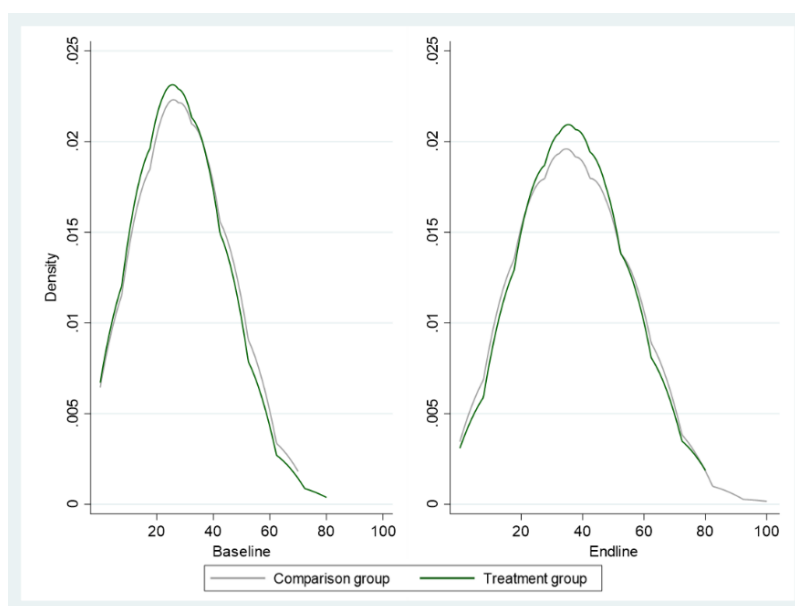
Figure 19 Distribution of nutrition knowledge scores of caregivers

66.   Although estimates show no impact on nutrition knowledge scores on the whole, heterogeneities in the effectiveness of the program were found in the subgroup analysis. As shown in Figure 20, for caregivers in the wealthiest quintile, the nutrition education increased the nutrition knowledge scores in the treatment group by 5.31 points. Detailed regression results can be found in Annex 9 (Table A9.24).
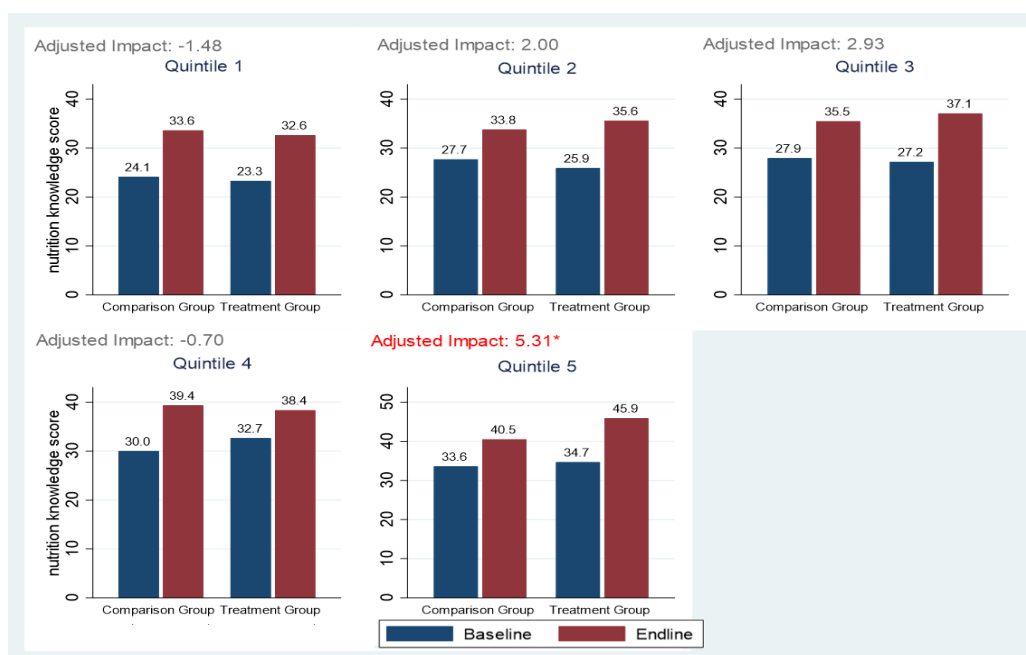
Figure 20 Heterogeneous program impact on nutrition knowledge scores of caregivers, by wealth index

67.   The above results imply that the program may not increase nutrition knowledge scores of caregivers as a whole, but they may be effective in increasing the scores of wealthier caregivers.

68.   There are two possible explanations. On the one hand, the nutrition education for caregivers organized by the preschools was inadequate. On the other hand, because the majority of primary caregivers are elderly and less educated, they might not understand or pay enough attention to the nutrition education.

Nutrition knowledge of preschool staff

69.   Preschool staff (principals, teachers, nurses and kitchen managers) participated in nutrition education about 4 times from 2018 to 2021, which was trained by nutritionist from Chinese Center for Disease Control and Prevention and experts from the Institute of Food and Nutrition Development of MARA on key knowledge about nutritional diets, cooking techniques and child feeding.

70.   As shown in Table 7, the score of the treatment group increased by 5.7 points from baseline to endline, while only 2.8 scores increased in the comparison group. And the score of the treatment group was significantly higher than that of the comparison group at endline. Although the average nutrition knowledge score of the treatment group was higher at baseline, the difference between two groups became greater after the intervention (from 5.5 to 8.4). Figure 21 shows the distribution of nutrition knowledge scores of preschool staff, in which the difference between two groups is more intuitively from baseline to endline.

Table 7 Nutrition knowledge scores of preschool staff

Group	Treatment group (1)	Comparison group (2)	P-value [H0: (1)=(2)] (3)
Baseline	52.3	46.9	0.0306
Endline	58.1	49.7	0.0327
Change from Baseline to Endline	5.7	2.8	-

71.   The results imply that nutrition education is effective in improving the nutrition knowledge for preschool staff. Due to the small sample size after matching caused by the high mobility of preschool staff, Difference in Difference method and heterogeneity analysis are not applicable here.
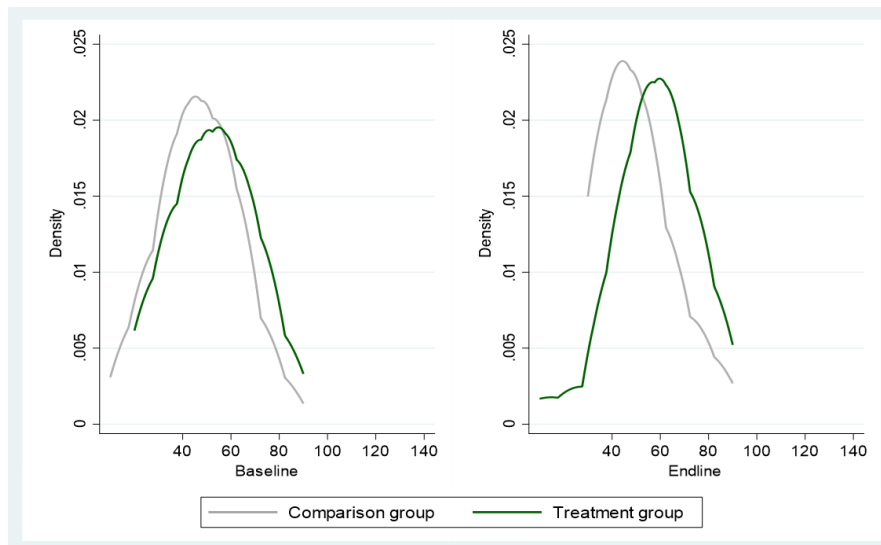
Figure 21 Distribution of nutrition knowledge scores of preschool staff

2.1.5. SUB-QUESTION 5: TO WHAT EXTENT DID THE NUTRITION PROGRAM IMPROVE THE FOOD CONSUMPTION AND NUTRIENT INTAKE OF CHILDREN?

Dietary Diversity

72.   It is widely accepted that dietary diversity is a key indicator of a healthy diet (Kennedy et al., 2007; Otsuka et al., 2016; Zhao et al., 2017). Scholars found that increased dietary diversity (as measured by the dietary diversity score) is strongly associated with micronutrient sufficiency and leads to positive health outcomes (Onyango et al., 1998). Literature showed that children with a more diverse diet had higher anthropometry measurements on average, including HAZ, WAZ, and WHZ (Frempong and Annim, 2017). On the other hand, a less diverse diet might increase the risk of stunting or underweight and even led to cognitive deficits in children (Black, 2003).

73.   Children's dietary consumption data were obtained by trained enumerators through face-to-face interviews. Questionnaires aimed to ask primary caregivers (mostly parents and grandparents) about what the children ate at home, as well as food eaten away from home on the most recent day from Monday to Thursday, the most recent Friday, and the most recent weekend. In order to be consistent with baseline data for the purpose of comparisons, we used the dietary consumption data on the most recent day from Monday to Thursday for the report purpose.

74.   Under the guidelines of the Food and Agriculture Organization (FAO) of the United Nations, we used Dietary Diversity Score (DDS) to measure the dietary diversity of children based on the dietary data described above. DDS was based on nine diverse food groups, including starchy staples, dark green leafy vegetables, other vitamin-A rich fruits, and vegetables, other fruits and vegetables, organ meat, meat and fish, eggs, legumes, nuts, and seeds, as well as milk and milk products (Details shown in Table A9.23 in Annex 9). The score was calculated by counting the number of food groups consumed on the most recent day from Monday to Thursday. Any individual food item in each food group consumed by children earned one point for the dietary diversity score, but different individual food items consumed in the same group would not be counted repeatedly. Therefore, the score range of DDS was from 0 to 9.

75.   Our data show that there was no significant difference in the change of DDS between treatment and comparison groups from the baseline to the endline surveys. Among children from the treatment group, the overall mean DDS increased from 4.39 in the baseline to 4.72 in the endline, while that in the comparison group increased from 4.34 to 4.65 during the same period (Figure 22). We then examined the differences in DDS by treatment status and found 0.31 in the treatment group and 0.33 in the comparison group, which was also statistically insignificant (Figure 23).
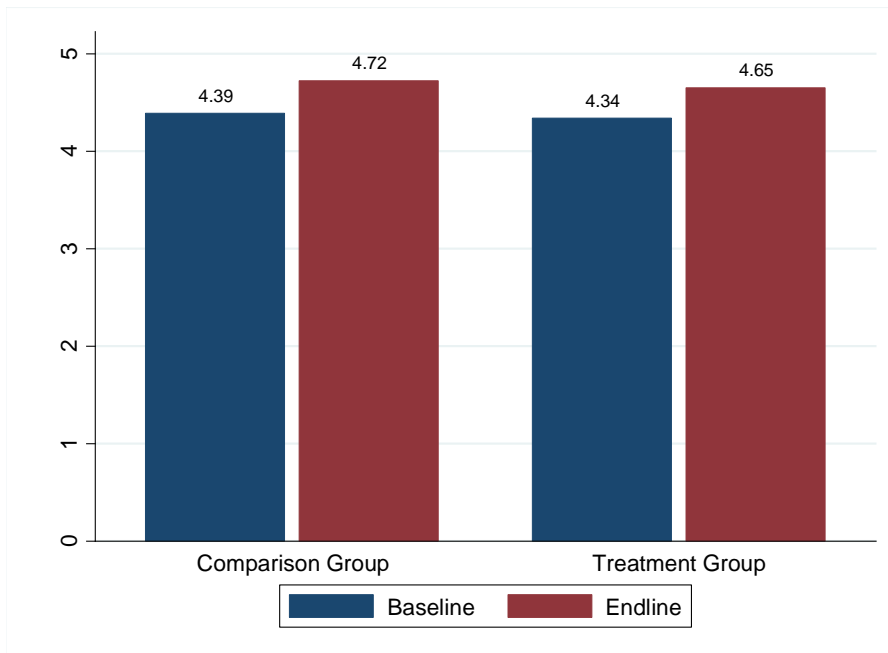
Figure 22 Average dietary diversity scores by treatment status over time
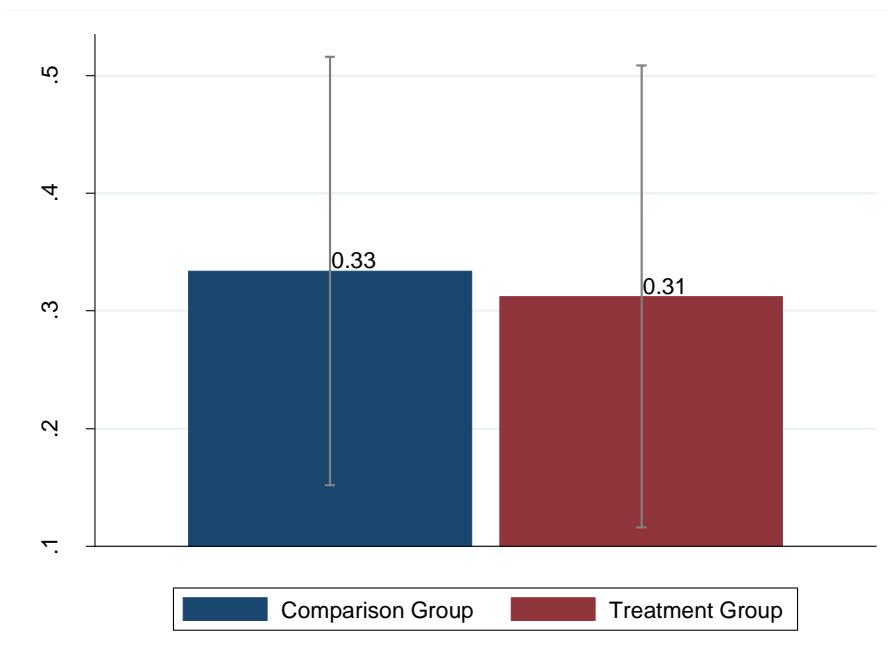
Figure 23 Differences of dietary diversity scores, by treatment status

76.   Caregivers' nutrition knowledge is one of the factors that affect children's at-home DDS. If caregivers' nutrition knowledge was improved, they would have been more likely to apply the acquired nutrition knowledge in children's daily lives and improve their daily feeding. However, results from caregivers' nutrition knowledge presented in the previous section showed that the program had no impact on caregivers' nutrition knowledge. Therefore, it was expected that the program had no impact on the DDS of children either.

77.   Another possible reason behind the lack of impact on children's DDS was that there might exist substitution or leakage. Anecdotes reported by some primary caregivers show that as children could enjoy better food nutrition intake in preschools, caregivers tended to pay less attention to their nutrition and diversity of food provided at home (this is what we call substitution). Moreover, some children brought food from preschools to share with other family members (usually siblings) at home (this is what we call leakage).

Such substitution or leakage might contribute to no significant improvement in children's DDS from the baseline to the endline. To capture whether these are an issue, the team asked the caregivers about children's food intake at home during the baseline and endline surveys. The results shown the diet quality of children in the treatment group had no significant differences compared to that of their counterparts, denying the existence of substitution. In addition, results failed to prove the existence of leakage given that few children (7.5%) in the treatment group brought food from preschools to home at the survey time.

78.   It should be noted that DDS at home was only part of the story. As the program aims to improve the children's food nutrition intake in preschool, the dietary and nutritional intake of provided school lunch should also be taken into account. The dietary intake for the school were analyzed using data from China Development Research Foundation (CDRF)'s digital technology and monitoring platform, where kitchen manager in each benefited preschool uploads key information of daily meals.

79.   For purposes of interpretation, the relative nutritional value of the school lunch is estimated according to whether it has met a Recommended Nutrient Intake, drawing upon 40% of a daily recommended nutrient intake standard. According to findings from CDRF, most of school lunches met the dietary intake standard, including cereals, meat, poultry, fish, soy products, and vegetables. The school meal also reached the standard nutritional intake of energy, protein, carbohydrates, iron, zinc, vitamin B1, vitamin B2, niacin, vitamin C and vitamin A. The only nutrient that did not meet the intake standard was calcium. In addition, excessive intake of sodium indicated excessive use of salt, therefore, the cooking method in the schools should be adjusted to reduce the use of salt.

Food Expenditure

80.   We also assessed program impact on children's monthly food expenditure as a proxy for household's investment in child nutrition. Food expenditure had been proved to be closely associated with the decreased risk of childhood malnutrition (Monsivais & Johnson, 2012; Kumar et al., 2019). If nutrition awareness of caregivers is improved by the school feeding program, household food expenditures on children are expected to increase, which can become an important driver for improving children's nutrition.

81.   We first displayed children's monthly total food expenditure[31] by treatment status in the baseline and endline. As shown in Figure 24, children's total food consumption in both groups had increased from nearly 460 yuan per month in the baseline survey to 590 yuan per month in the endline survey. However, as shown in Figure 25, no significant differences between the two groups were found in the changes in children's monthly total food expenditure from baseline to endline.

---

[31] Monthly food expenditure in the dataset was winsorised at the 2.5th percentile and 97.5th percentile to set extreme outliers equal to less extreme values.
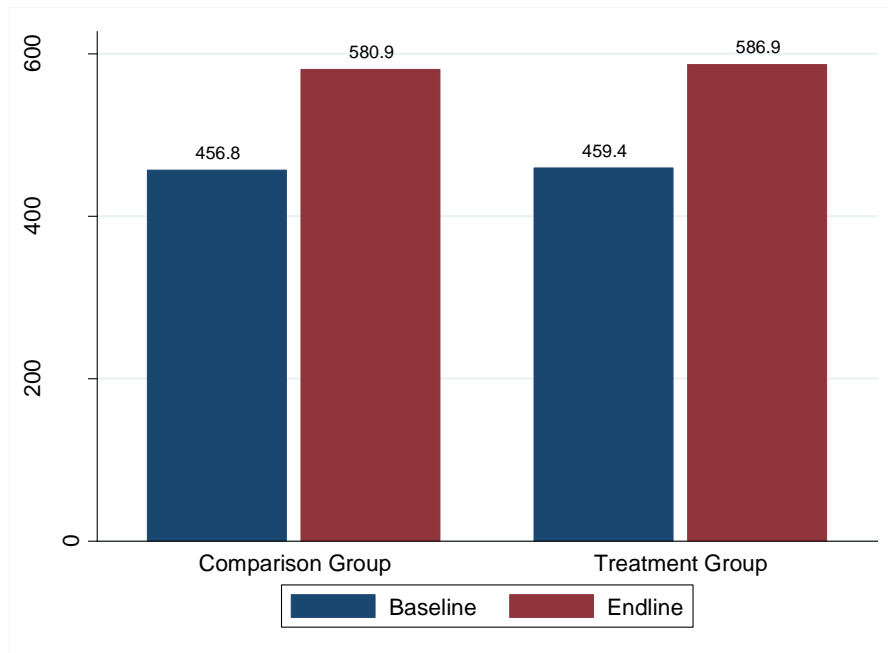
Figure 24 Children's total food consumption by treatment assignment status over time, yuan per month
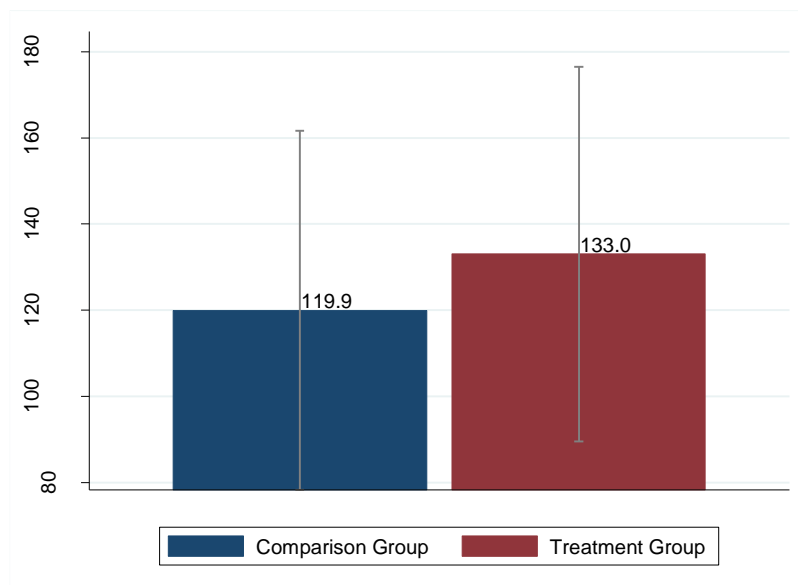
Figure 25 Differences of children's total food consumption by treatment assignment status, yuan per month

82.   Then we showed food expenditure patterns on specific food groups by treatment status. We mainly discussed three types of food: 1) Snacks including candies, cakes and beverage, which was highly associated with weight gain and obesity risks (Drewnowski & Specter, 2004; Malik et al., 2006); 2) Fruits, the consumption of which helped to reduce early childhood malnutrition (Campbell et al, 2008); and 3) Meat, eggs and milk productions, which were typical animal source foods supplying high-quality and readily digested protein, energy and micronutrients (Neumann et al., 2002). The results shown in Figure 6.26 indicated that children's food expenditure on the three types of food all increased from baseline to endline. However, we could not find any significant differences between the treatment and comparison groups in any types of food expenditure changes in Figure 27.
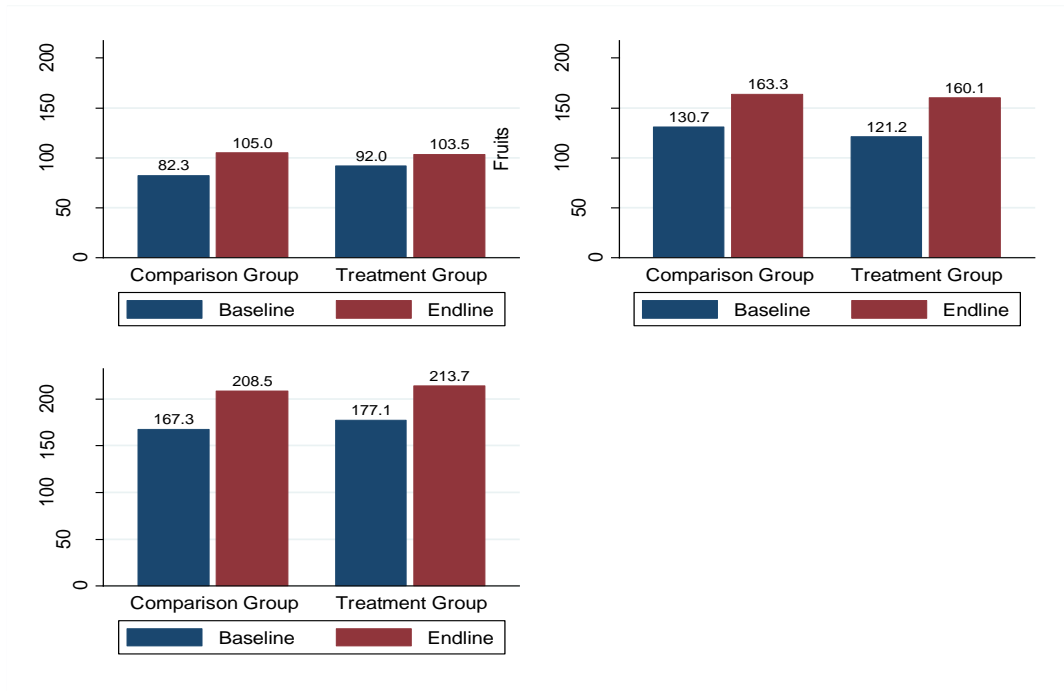
Figure 26 Children's different types of food consumption by treatment assignment status over time, yuan per month
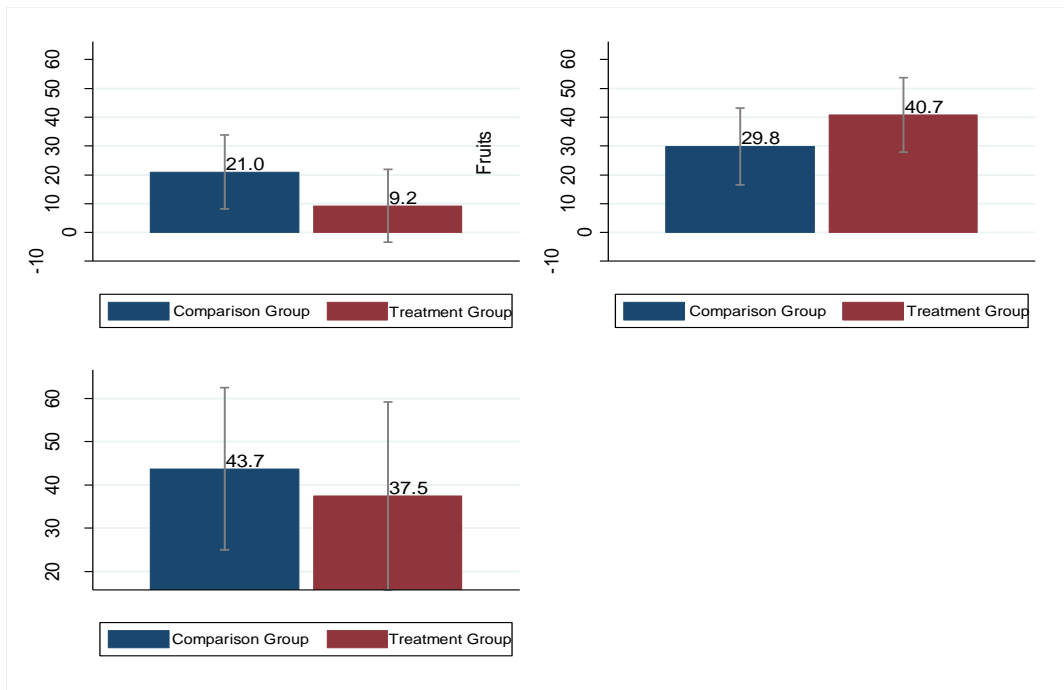
Figure 27 Differences of children's different types of food consumption by treatment assignment status, yuan per month

School Meal Quality

83.　　In the program design, children's nutrition and developmental outcomes were also expected to be improved by providing nutrient-rich school meals. We focused on the changes of school meal frequency and quality in sample preschools[32] from baseline to endline, and compared them by treatment status.

84.　　Figure 28 shows the types of school meals provided by sample preschools from baseline to endline by treatment status. It indicated that school lunch was always available in the sample preschools while the main changes were the supply of breakfast and snacks. In the baseline, relatively fewer preschools in the treatment group than in the comparison group provided breakfast (70% VS 75%). During the intervention period, 10% of the treatment preschools began to supply breakfast and 50% of them began to supply snacks. Similarly, the proportion of preschools that provided snacks also increased in the comparison group by 33%, although 17% of them stopped providing breakfast. Overall, the frequency of school meal increased in both groups since more preschools began to provide breakfast and snacks during the program.
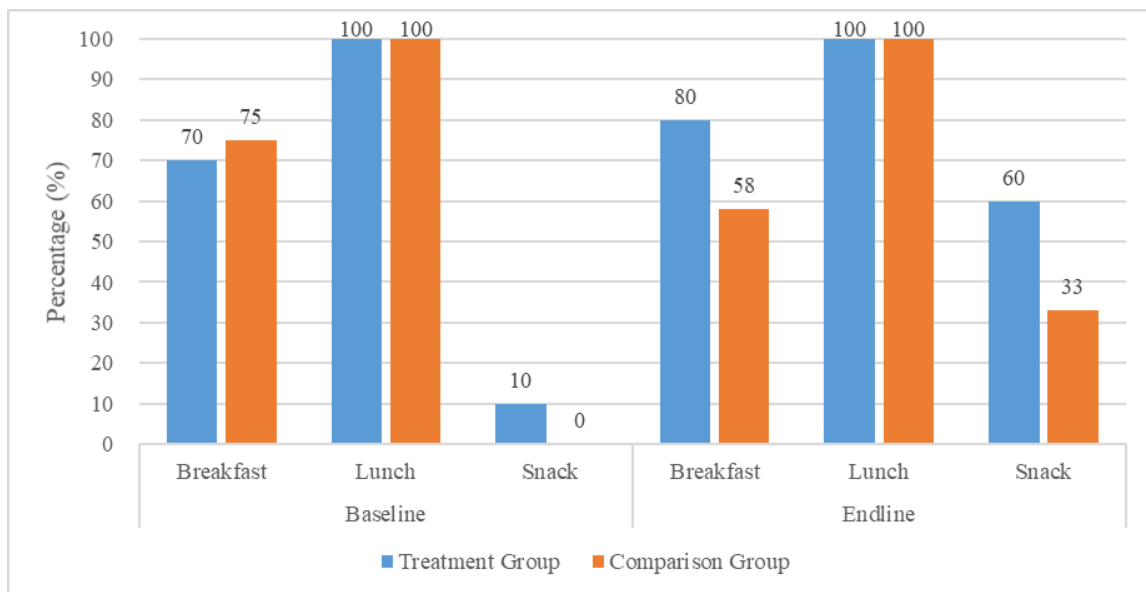
Figure 28 Types of school meal provision in sample preschool over time, by treatment status

85.　　However, the increase of school meal frequency was not equal to the improvement of dietary quality. We used diet records in a typical school day provided by kitchen managers, which recorded all food consumed in each preschool. Following the method in the caregiver module, we calculate dietary diversity scores (hereinafter referred to as DDS) to proxy the quality of school meals. Figure 29 showed the changes of DDS for school meals from baseline to endline by treatment status. As expected, DDS for school meals in the treatment group increased from 4.8 to 6.7, indicating each treatment preschool on average provided nearly two more types of food. However, the comparison group also increased in the quality of school meals, with 2.4 higher scores of DDS. When the changes of DDS for school meals were compared between groups, shown in Figure 30, the group difference was statistically insignificant. Therefore, the findings could not provide evidence that school meal quality in the treatment group had been improved more than that in the comparison group.

---

[32] Ideally, we would analyze frequency and quality of school meals in 26 sample preschools. However, with two preschools stopping running before the endline survey and two preschools refusing to provide related information, we specified the sample to 22 preschools with complete and two-wave information of school meal frequency and quality.
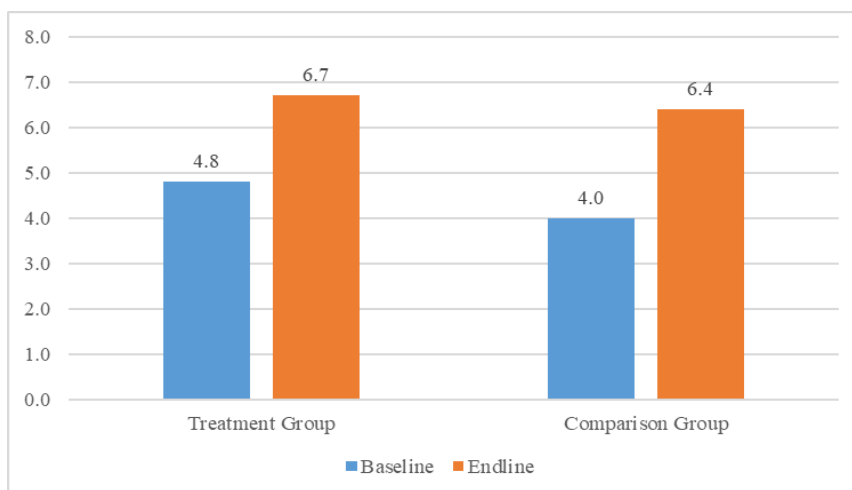
Figure 29 DDS for school meals over time, by treatment status
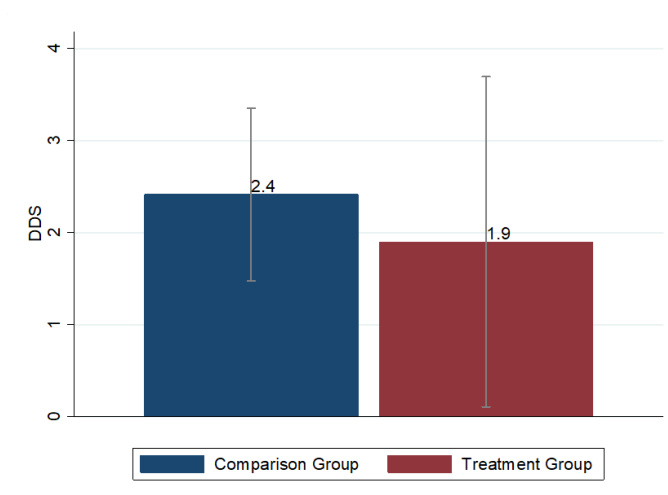
Figure 30 Changes of DDS for school meals, by treatment status

2.1.6. SUB-QUESTION 6: TO WHAT EXTENT DID THE NUTRITION PROGRAM IMPROVE THE INTERACTIVE PARENTING OF CAREGIVERS?

86.   The school feeding program was also expected to improve children's development by promoting interactive parenting practices. In this section, we discussed the practice of interactive parenting by treatment status.

87.   We focused on three interactive parenting practices: playing with the child, telling stories to the child and singing to the child on the day before the survey. These three indicators were chosen because the linkage between them and child development had been proved in psychological and biological literature (e.g., Luo et al., 2019). Figure 31 shows the shares of caregivers who told stories with the child, sang songs with the child and played with the child respectively by treatment status. Our data did not provide any evidence supporting that the school feeding program could improve caregivers' engagement in interactive parenting. Specifically, the shares of caregivers who practiced these activities are not significantly different between the treatment and the comparison groups. The shares of caregivers in both groups who played games with the child yesterday are 41%. Caregivers were generally less likely to tell stories or sing songs with the child. Caregivers who told stories and sang songs accounted for 13% and 17% respectively in the treatment group, while 11% and 15% in the comparison group.
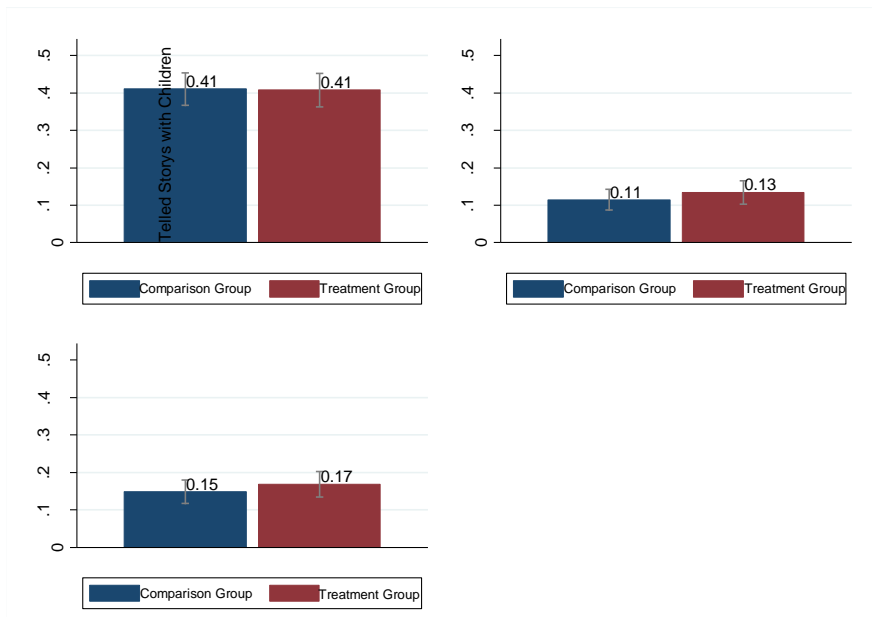
Figure 31 Interactive parenting by treatment assignment status

88. Three additional indicators related to parent-child interaction were used to double-check the school feeding program impact on interactive parenting. The results did not lend any support that the school feeding program would improve interactive parenting patterns, either. The first one is the frequency that caregivers showed affection to the child, exemplified by hugs and kisses. Figure 32 shows that the child averagely receives at least one kiss or hug displayed by the caregivers. However, no significant difference is found between the treatment group and the comparison group. The next two indicators are the daily time the child spent playing alone and spent watching TV or videos (collectively referred to as screen time). These activities potentially squeeze out times available for parent-child interaction, therefore, the higher the indicators mean the worse performance of interactive parenting. Figure 33 shows that children in the comparison group spent 41 minutes per day playing alone and 86 minutes as their screen time, and their counterparts spent respectively 49 and 85 minutes. However, insignificant differences in daily play alone time and screen time of the child between the treatment group and the comparison group were found.
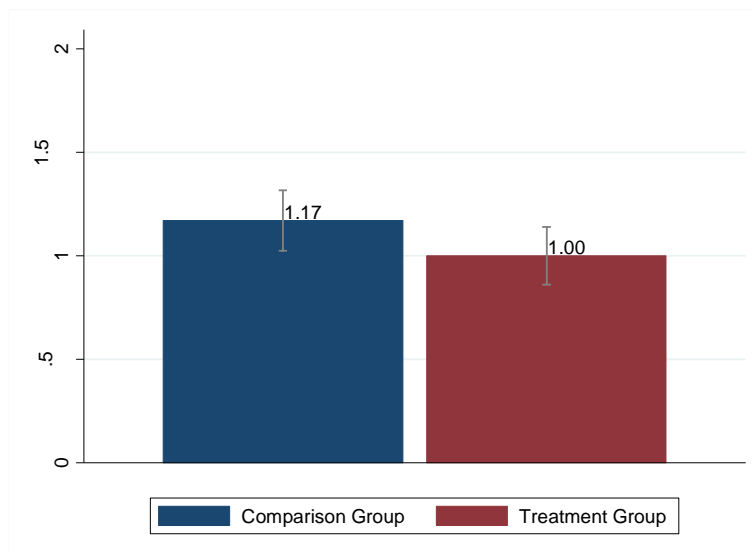
Figure 32 Times that caregivers showed love to the child yesterday by treatment status
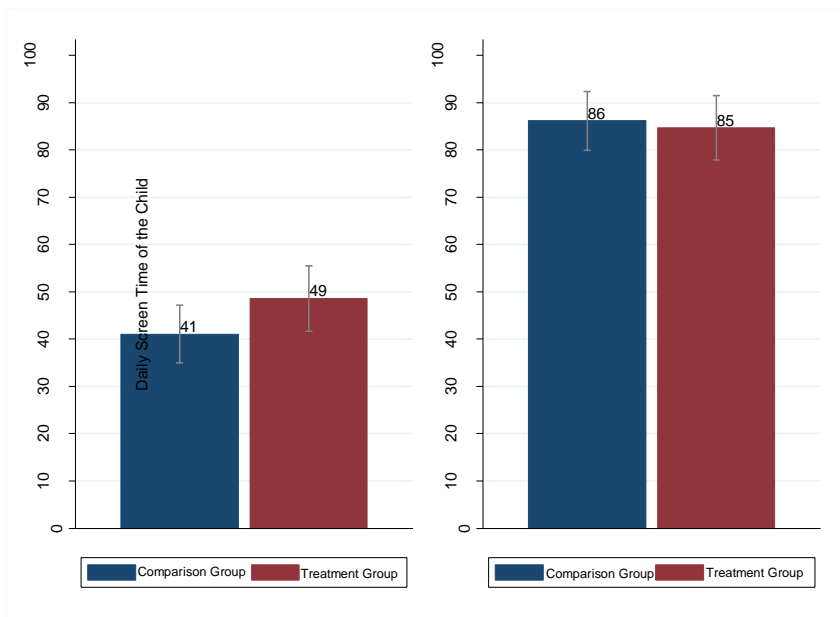
Figure 33 Daily play alone time and screen time of the child by treatment status

89.   Since the evaluation could not provide clear evidence that school feeding program improves average children's nutrition, cognition and social emotions, we examined three critical elements underlying the link between school feeding program and target outcomes, including nutrition knowledge score, feeding practice and interactive parenting.

- The school feeding program improved the nutrition knowledge scores of preschool personnel but not that of caregivers. More targeted and effective nutrition education methods (behavioral change communication, preferably) are required to improve the effects of nutrition training.

- The dietary quality at home and in preschools were both improved in both groups similarly. In other words, the school feeding program had little impact on children's dietary quality.

- Interactive parenting quality was similar between groups. The results were reasonable given that the nutrition training, as well as other program activities, targeted more on nutritional purposes.

90.   However, not all these three underlying mechanisms came out significant, indicating that the impacts of school feeding program on children's nutritional, cognitive and socio-emotional outcomes could not be realized by these mechanisms. One possible explanation could be that due to the COVID-19 or challenges sticking with the intervention protocols, there might be some spillover effects from the treatment group to the comparison group as the latter group could also benefit from improved nutrition knowledge, better school and home feeding practices, and interactive parenting behaviors. The findings help to explain why few significant impacts were found in the program evaluation process, pointing out the importance to discuss the efficiency and appropriateness of the program implementation.

2.2. EVALUATION QUESTION 2: WHAT IS THE IMPACT OF THE NUTRITION PROGRAM ON LOCAL SMALLHOLDERS?

2.2.1. SUB-QUESTION 1: TO WHAT EXTENT DOES THE NUTRITION PROGRAM BOOST AGRICULTURAL PRODUCTION OF SMALLHOLDERS?

92.   When we compare the agricultural production of smallholders in the treatment and comparison groups, we conducted two types of comparisons. Comparison between treatment group and comparison group in the endline survey, and comparison of the changes between treatment group and comparison group from baseline to endline surveys.

93.   Data from the endline survey show that treatment group is more active in agricultural production in terms of land use, crop planting, livestock raising and agricultural technical training.

Land use

94.   The area of cultivated land (including paddy field and dry land) of treatment group is significantly larger than that of the comparison group. Moreover, the area of rented-in dry land in the treatment group is significantly larger than the comparison group. These observations seemed to indicate that the program helped smallholders expand their agricultural activities, as expected from the project design. In addition, the numbers of plots of both paddy fields and dry land in treatment group is significantly bigger than those of the comparison group.

95.   When we compare the changes from baseline to endline surveys, the picture is different. The area of cultivated land of treatment group is slightly reduced while that of the comparison group increased. The reason is the area of dry land of treatment group reduced although the area of paddy field increased. A significant increase in the rented in paddy field contributed to the increase in the paddy area of the treatment group.

Application of chemical fertilizer and pesticide

96.   Aside from contracting, another component of the intervention is to train farmers to enhance their awareness of green production and provide them with organic fertilizer for free. Compared with baseline data, both groups increased application of chemical fertilizer in vegetable production but the increase in the treatment group is smaller than that of the comparison group. Both groups decreased application of organic fertilizer but the decrease in the treatment group is much smaller than that of the comparison group.

97.   Data from the endline survey show that the application of chemical fertilizer and pesticide for rice production was much lower in the treatment group than in the comparison group, indicating positive impact of the training on green production.

Table 8 Comparison of agricultural production between the Treatment and Comparison Groups in the endline survey

	Endline sample	Treatment Group	Comparison Group	Difference in means
Observations	106	58	48	
Land Use				
Area of Paddy field (Mu)	3.158	3.871	2.296	1.575*
No. of Plots of Paddy field	3.160	3.672	2.542	1.131*
Area of Dry land (Mu)	4.418	5.231	3.435	1.796**
No. of Plots of Dry land	5.019	5.759	4.125	1.634*
Area of Transferred Dry land (Mu)	0.746	1.140	0.271	0.869**
Crops				
Chemical fertilizer used for rice (Jin/Mu)	57.893	42.025	77.067	1.352***
Organic fertilizer used for rice (Jin/Mu)	25.211	21.322	29.911	3.733***
Pesticides used for rice (g/Mu)	153.341	127.778	184.229	-35.041*

Agricultural technology training

98.　　The data showed that between May 2020 and April 2021, the participation in agricultural technology training was statistically the same in the treatment and comparison groups. This result held regardless of the measures of participation, being the share of smallholders who participated in trainings, the number of participants, the number of crops involved, or the number of techniques covered in the training.

99.　　The data also showed that the share of households who participated in training did not increase in the treatment group yet it did increase in the comparison group. The number of trainings received in both groups decreased. One possible explanation is the impact of the COVID-19 as reflected from local program officers.

Table 9 Agricultural technology training

Training	Endline sample	Treatment	Comparison	Diff
Share of smallholders participated in agricultural technical training	26.4%	31.0%	20.8%	10.2%
Average number of participations	0.302	0.379	0.208	0.171
How many types of crops were involved in these training sessions?	0.311	0.379	0.229	0.150
How many types of techniques were involved in these training sessions?	0.377	0.431	0.313	0.119

2.2.2. SUB-QUESTION 2: WHAT IS THE IMPACT OF NUTRITION PROGRAM ON PRODUCTION DIVERSIFICATION OF SMALLHOLDERS?

Crops production

100.　　One objective of this program is to increase the crop diversity through providing free agricultural inputs. Results from the endline survey showed that households in the treatment group planted more crop varieties than those in the comparison group. This is especially true when it comes to vegetables. As it turned out, the vegetable acreage of the treatment group was 4.32 mu (or 0.29 hectare), more than 7 times that of the comparison group.

101.　　Similar result is shown from the changes of the two groups in the number of varieties planted. The number of varieties planted by both groups of households increased. Specifically, the treatment group increased from 2.88 to 6.63 whereas and the comparison group from 2.12 to 5.64. The pattern is more obvious when it comes to the number of vegetable varieties planted, which increased from 1.13 to 5.00 for the treatment group whereas from 0.72 to 4.32 for the comparison group. It is worth noting that the share of households in the treatment group that planted vegetables increased from 50% in the baseline to 94% in the endline. By contrast, this number in the comparison group increased from 32% to 40% during the same period. In terms of the sown area of vegetables per household, it increased by 2.12 mu in the treatment group whereas by 0.79 mu in the comparison group. Surprisingly, the yield of vegetables decreased in both groups but the decrease in the treatment group is much smaller than in the comparison group. Field interviews show that one possible reason for the yield decrease is the heavy rain.

Livestock diversity

102.　　Similar to what were found on crop production, the endline survey data also showed that the number of livestock varieties raised by treatment group household was significantly higher than that of the comparison group. Moreover, the number of the two most popular animals (Chicken and Pig) raised by smallholders was also larger in treatment group than in the comparison l group.

Table 10 Production diversity

Crops diversity				
No. of varieties planted	7.217	9.052	5.000	4.052***
Area of cultivated vegetable (Mu)	2.631	4.321	0.589	66.013**
Livestock diversity				
No. of varieties raised	1.491	1.897	1.000	0.897***
No. Full grown Pig	0.302	0.448	0.125	0.323**
No. Full grown Pig	20.755	23.241	17.750	8.192*

2.2.3. SUB-QUESTION 3: TO WHAT EXTENT DOES THE NUTRITION PROGRAM CONTRIBUTE TO LOCAL SMALLHOLDERS' INCOME?

103. Smallholders in the treatment group has a higher total gross income level than their peers in the comparison group. Similar to the findings from the baseline survey, government transfer income still accounted for the largest part of their income sources. This is due to the fact that 84% of the households in the endline survey are previous registered poor households.

Table 11 Gross household income and income sources (endline)

	Treatment group	Comparison group	Difference in means
	(1)	(2)	(3) = (1) - (2)
Gross household income	73,087	52,585	20501**
a) Wage income	13,116	13,920	-804
b) Business income	1,132	2,266	-1134
c) Property income	79	1,381	-1301
d) Transfer income	58,759	35,016	23,742**

104. The table below presents a difference-in-difference comparison of the gross income per capita for the 37 households that participated in both the baseline and endline surveys. Relative to comparison households, treatment households see their gross income per capita increase by RMB 689.65 more from the baseline to the endline surveys, or approximately 10.5% more.

Table 12 Comparison of income between treatment and comparison group

Gross income per capita	Treatment group	Comparison group	
	(1)	(2)	(1)-(2)
Baseline	3632.88	8034.64	-4401.75
Endline	6570.02	10282.12	-3712.10
Endline-Baseline	2937.14	2247.48	689.65

Case study

105. Based on the focus group discussions, we observed that contracting smallholders enjoyed multiple benefits delivered by the pilot. First, food loss and waste of farmers had been reduced. Prior to the intervention, surplus vegetables and fruits often got rotten in the field, causing food loss. After the agreement was signed, surplus agricultural products could be sold to preschools at market prices to earn some income. Farmers reported that this part of income were used as pocket money for daily household expenses. At the same time, compared to selling on the market, smallholders did not have to wait in the market when they sell to preschools, which saved time, and reduced transaction costs to a large extent. In addition, prices could be volatile when selling on market while prices paid by preschools were relatively stable.

2.2.4. SUB-QUESTION 4: WHAT IS THE IMPACT OF THE NUTRITION PROGRAM ON DIETARY DIVERSITY OF SMALLHOLDERS?

106. The endline data showed that smallholders in the treatment group had a higher dietary diversity score (DDS) than those in the comparison group. Figure 35 also showed difference in the proportion of

households consuming each of the nine food groups by treatment status. Treatment group had a higher probability of consuming all nine food groups except for meat and cereal, which they consumed every day during the recalled period.
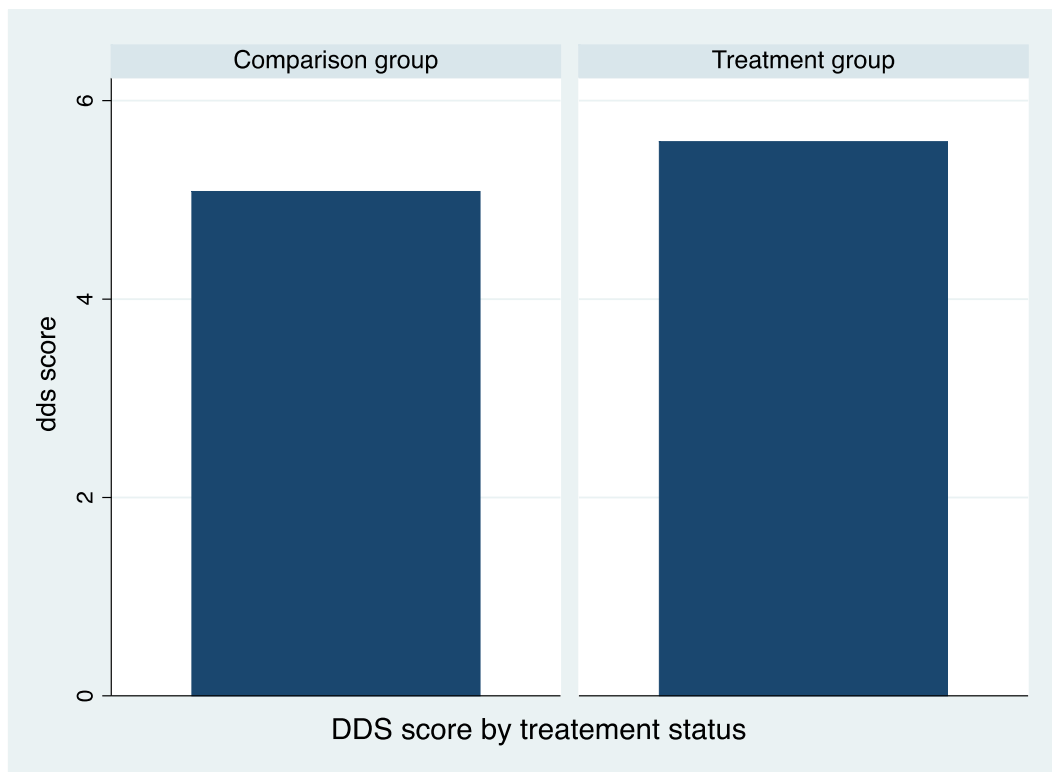
Figure 34 Dietary diversity score of farmer households by treatment status in the endline survey
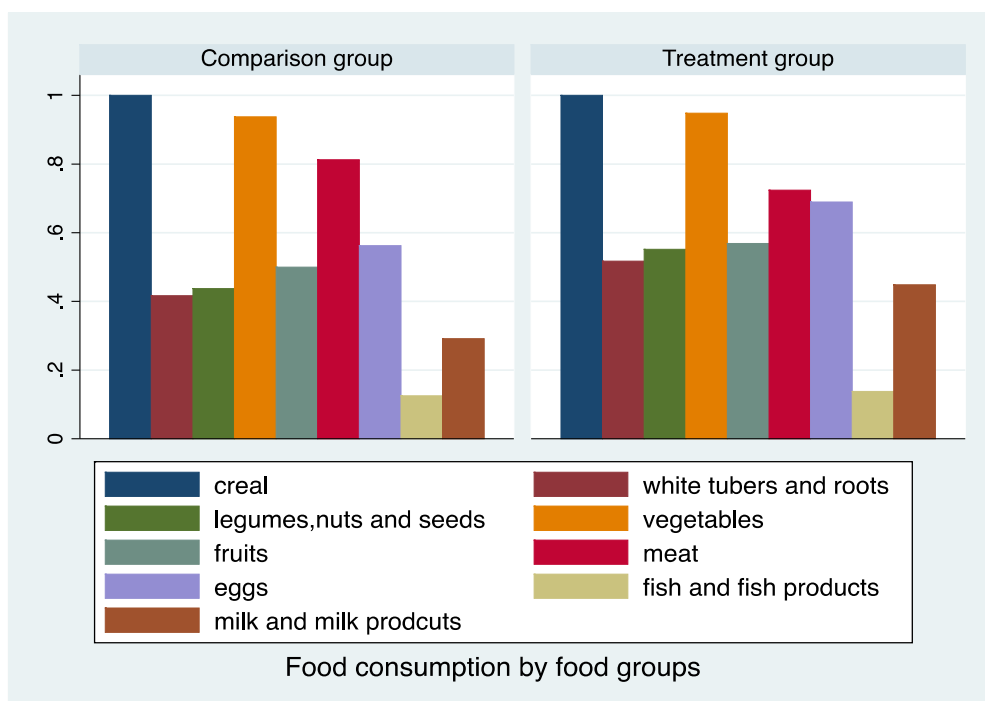
Figure 35 Individual food group consumption of farmer households by treatment status in the endline survey

107.   **Case 1.** Xiang Mei, female, 54 years old, from Duishan Village, Duishan Township, Yongshun County, who was from registered poor household. A grandson of hers attended the Smurf Preschool in the same

Township. After signing the "Agricultural Product Planting, Production and Marketing Agreement" with the Smurf Preschool, she was employed by the preschool as a chef. Her household grew rice and a variety of vegetables, and more than half of the output was supplied to preschools. As a preschool chef, she was familiar with the preschool's recipes, and understood the daily ingredients of the preschool and the long-term needs of the types of agricultural products. Xiang Mei fresh-picked the corresponding agriculture products directly from her farm according to the preschool's recipes and brought them to the preschool every day. It not only guaranteed that children's diet was in strict accordance with the recipe, but also utilized the freshest vegetables to ensure food safety. After receiving the nutritional knowledge training provided by the Xiangxi Prefecture government, the Smurf preschool's diet has undergone some great improvements. The children's dietary diversity has increased. The preschool advised her on what types of agricultural products to plant in the field in the next season based on the preschool's demand for ingredients. Xiang Mei was also willing to change the planting types, and has increased the planting of peanuts, pumpkins and other varieties.

108.   **Case 2.** Liu Guilian, female, from Xiachi Village, Shipai Town, Longshan County. There are no young and middle-aged labors in her family, and they mainly live on government subsidies and wages from part-time jobs. They grow a small amount of vegetables for preschools. The production scale is small, and the seeds distributed by the project office free of charge can fully meet their planting needs. Before signing the agreement, only corn, sweet potatoes and other agricultural products were planted all year round. After the agreement was signed, all the seeds issued by the project office were planted in the fields. There was no need to buy other seeds by herself, and a variety of agricultural products were added.

109.   We also compared the Nutrition Knowledge Score (NKS) between the two groups of households. Our data showed that the NKS in treatment group increased from 28.13 to 33.75 while that of comparison group decreased from 28.80 to 24.80.
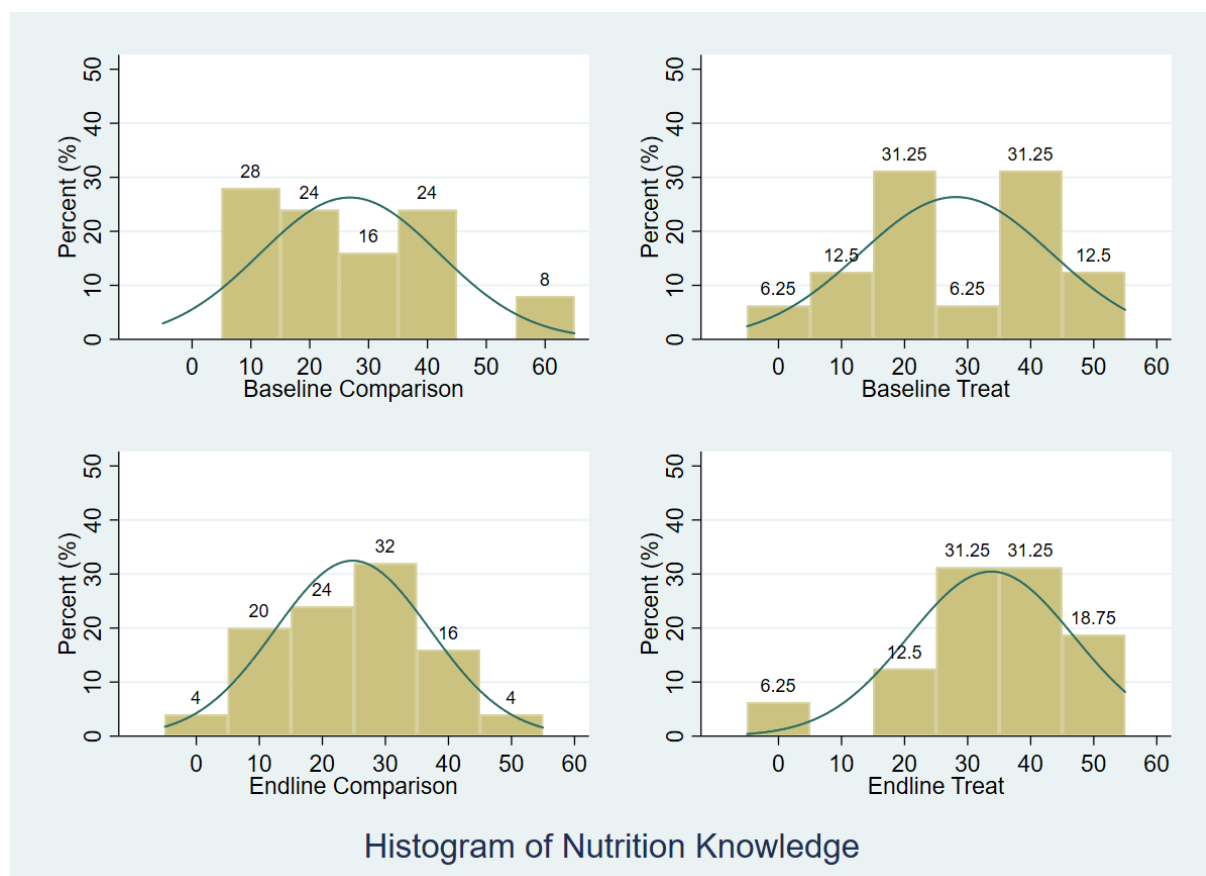
Figure 36 Comparison of Nutrition Knowledge between Treatment and Comparison Groups

2.2.5. SUMMARY

110.   This section analysed the smallholder direct food supply programme implemented in Xiangxi, China. The programme was designed and implemented as an integral part of the WFP free meal programme. Using

field survey, semi-structured key informant interviews and focus group studies, it analysed the impact of the nutrition program on local smallholders specially on agricultural production, marketing channels of the production, local smallholders' income and food security and nutrition, and uncovered both good practices and challenges of the direct purchase model implemented in Xiangxi.

111.   It is worth mentioning that the targeted beneficiaries of the smallholders varied in endline survey from the baseline survey as one of the criteria of choosing the targeted beneficiaries is smallholders having a child attending the preschool with which they are about to sign a contract. This indicates that as the child graduated from preschool, the householder will be no longer qualified. This causes some attrition although this criterion serves and proves to be an effective measure to ensure the food provided meet the food safety and quality standard. Hence both comparison analysis of the treatment group and comparison group in endline survey and the changes of the key indicators of householders both in baseline survey and endline survey were analysed.

112.   The results showed that the program contributed to local smallholders' vegetable production diversity, income, nutrition knowledge and dietary diversity. Treatment group is more active in agricultural production in terms of land use, crop planting, livestock raising and agricultural technical training. The rented area of paddy field increased significantly by smallholders in treatment group. The number of varieties planted by treatment households increased more than comparison group especially in vegetables. Similar as livestock production. The share of smallholders participated in agricultural technical training in treatment group is higher than in comparison group yet impacted by the COVID-19, trainings organized by the program office were much limited as reflected by smallholders.

113.   The direct purchase model enabled contracted smallholders selling agricultural products to preschools which ranks the third major marketing channel after self-consumption and feed stuff while for comparison it is to sell to individuals. Besides the benefit of selling the products at market price, this direct purchase also saves the smallholders' waiting time and transportation to sell in market. An approximately 10.5% increase of the gross income per capita of the treatment householders was found from the difference-in-difference analysis.

114.   It is positive to find out that smallholders in the treatment group had a higher dietary diversity score (DDS) than those in the comparison group although the DDS increased more in comparison group. It is acceptable considering the higher the DDS is, the more difficult to increase the diet diversity. Also food loss and waste was reflected to be reduced and the Nutrition Knowledge Score (NKS) increased significantly in treatment group compared with comparison group.

115.   This analysis shows that it is beneficial to build a local, inclusive, sustainable and resilient food system. The identification of the beneficiaries especially those smallholders whose child is in the preschool is proved to be a good practice. Despite of the benefits, this study also identified several constraints for sustaining or scaling up the current programme, including the unstable supply under the farm to preschool model, the insufficient supply-side intervention, the transportation constraints and the expected decrease of demand from the local preschools. Due to the outbreak and long lasting COVID-19, a few interventions such as training is delayed, and together with some negative impact of whether shocks, the agricultural production was not significantly improved.

116.   These results imply that in order to improve the effect of the program, continuous supply side interventions are needed to stimulate smallholder's participation in the school feeding market and beyond. Moreover, to further increase the impacts of the demand-assisted approach, the structured demand could be extended to the broader public procurement, such as public demand from local governments and primary schools, given the limited demand size of the preschools. Providing more job vacancies such as preschool caterers and cooks may also be an avenue to poverty reduction and income increase of local smallholders.

2.3. EVALUATION QUESTION 3: HOW APPROPRIATE IS THE PROGRAM IMPLEMENTATION?

2.3.1. SUB-QUESTION 1: TO WHAT EXTENT ARE THE PROGRAM'S DESIGNED OBJECTIVES, TARGETING STRATEGIES, AND ACTIVITIES CONSISTENT WITH RELEVANT NATIONAL POLICIES, STRATEGIES, AND GUIDANCE ON NUTRITION, CHILD DEVELOPMENT, AND POVERTY REDUCTION?

117.   The program is well aligned with China's national priorities, including nutrition, child development and poverty reduction policies. The nutrition program was designed at a time when there is a growing realization in China's policy circles that economic growth alone cannot reduce absolute poverty and inequality, that investments in human capital are needed to sustain growth, break the intergenerational transmission of poverty, and improve social cohesion. One critical step to form a successful human development strategy is to invest in early childhood development (ECD): nutrition, care, education, and development for children from birth to six years of age. China's ECD services are guided by the Law on Protection of Minors approved by the People's Congress and by regulations and guidelines set forth by the State Council.

118.   The program is particularly aligned with the Government's emphasis on reducing child malnutrition. The National Nutrition Plan (2017-2030) released by the Chinese Government sets goals to reduce stunting among children under 5 to less than 7% and reducing anemia prevalence to less than 12% by 2020. It also complements other Government-funded initiatives. Present in China, infants of 6-24 months are provided with complementary micronutrients by the "nutrition package" for the "First 1,000 days of life" program promoted countrywide, and 6- to 15-year-old rural area students receiving compulsory education are benefiting from comprehensive nutrition improvement programs. However, the Government has not yet developed any nutrition improvement program for the 3- to 5-year-old preschool children. Therefore, the program has been instrumental in addressing this "gap" of nutrition intervention and raising the profile of preschool child nutrition on the national agenda.

119.   The program is also closely aligned with China's poverty reduction strategies. Considering that there are regional and rural-urban disparities in childhood development, more efforts are directed to western and rural regions, especially those in the poverty-stricken areas, in line with the country's anti-poverty movement. A series of mutually reinforcing and cross-referenced guidelines on human development and poverty reduction, with focus on ECD, were promulgated in support of this policy focus. The 13th Five-Year Plan for Poverty Alleviation, the Three-Year Guidelines of Winning the Battle against the Poverty, and the National Development Plan for Children in Poor Areas (2014-2020) all highlight the necessity to improve child health and nutrition in poor counties. Nutrition education campaigns are also viewed as key strategies to raise the awareness of caregivers on nutrition and encourage good child feeding practices. The program's integrated interventions on school feeding and nutrition education in national-level poor counties of Xiangxi are in line with the ECD strategy advocated by the Chinese government. The program element on boosting agriculture production of smallholders and increasing their incomes also relates to the government's interests on poverty reduction and rural transformation

2.3.2. SUB-QUESTION 2: TO WHAT EXTENT ARE THE PROGRAM'S DESIGNED OBJECTIVES, TARGETING STRATEGIES, AND ACTIVITIES CONSISTENT WITH UN AND WFP POLICIES, STRATEGIES, AND GUIDANCES ON NUTRITION, CHILD DEVELOPMENT, SCHOOL FEEDING, AND POVERTY REDUCTION?

120.   At the UN level, the nutrition program is in closely line with the Sustainable Development Goal 2 "end hunger, achieve food security and improved nutrition, and promote sustainable agriculture" in the 2030 Agenda. The nutrition program also falls into the priority area of poverty reduction and equitable development in The United Nations Development Assistance Framework (UNDAF) for the People's Republic of China. Equitable distribution of and access to social support such as food, education, and health care, with focus on the most disadvantaged and vulnerable populations, is highlighted as the UN's efforts to support China to promote innovative poverty alleviation mechanisms and develop a more equitable and resilient society.

121.   The program is closely related to WFP policies and strategies. Access to safe, nutritious, and sufficient food and consumption of nutritionally adequate diets with complementary actions to support utilization, lie at the core of WFP's mandate. In the newly published WFP School Feeding Strategy 2020 – 2030, WFP lays out its vision of working with governments and partners to jointly ensure that all primary school children have

access to good quality meals in school, accompanied by a broader integrated package of health and nutrition services. Although the strategy was designed with a focus on primary school children, the strategy explicitly expects a need in evidence on differences in nutritional needs for various age groups and sexes to tailor supports for countries where assistance is also provided to preschool children. This program is an opportunity to generate such evidence.

122. WFP School Feeding Strategy stresses that for all age groups school feeding needs to be paired with social and behaviour change communication to change food and nutrition behaviour. Nutrition education is an integrated approach in this program. WFP also encourages and facilitates combination of various Social and behavior change communication (SBCC) supportive methods, i.e. through advocacy (KOL support in social media), mass communication (posters and brochures disseminated which were developed with support from IFND), interpersonal communication (teacher-parent meetings), community mobilization (Meituan cooking demonstratio) etc.

123. As required by the WFP Corporate Partnership Strategy (2014–2017), various contributors are integrated into the nutrition program. WFP works with different levels of governments to deliver the program, generally in line with China's national strategies in terms of nutrition improvement, child development, and poverty reduction. Private sectors are involved to provide complementary support.

124. The WFP Strategic Plan (2017–2021) includes "improve nutrition" as one of its five strategic objectives. WFP's China Country Strategic Plan further clarifies that child nutrition improvement is the priority among its nutrition work and set the agency's first strategic outcome as "children living in targeted poverty counties have malnutrition rates reduced in line with national norms by 2020". To this end, the nutrition program designs a preventive approach to malnutrition, focusing on facilitating access to nutritious diets required by vulnerable children.

125. WFP's support for poor smallholders is listed as strategic outcome both in the WFP Strategic Plan (2017–2021) and China Country Strategic Plan. One of its experimented approaches is catalytic market creation activities through which program can transform its local procurement into a tool for improving smallholders' productivity, crop quality, and sales.

126. The importance of social protection is emphasized in relevant WFP policies but is not clearly presented in the nutrition program documents and implementation. The WFP School Feeding Policy suggests that school feeding within a social protection framework can act as a stable and reliable income transfer to poorer families, offsets education and food costs, provides important nourishment for children in chronically food insecure families, and decreases the risk that children are withdrawn from school. WFP's support to social protection and safety net programs also aims to strengthen the capacity of national systems, to forge linkages with the national food and health sectors.

2.3.3. SUB-QUESTION 3: IS THE PROGRAM DESIGNED AND IMPLEMENTED BASED ON A SOUND GENDER ANALYSIS THAT IS IN LINE WITH NATIONAL, UN, AND WFP GENDER POLICIES, CONSIDERING THE DISTINCT NEEDS AND PARTICIPATION OF BOYS/MALES AND GIRLS/FEMALES?

127. The design and implementation of the program are partially in line with gender policies. The program design is in line with WFP's Gender Policy (2015-2020) which confirms a shift from a women-focused to a gender-centered approach and calls for broader analysis of the challenges and opportunities in the lives of women, men, girls, and boys, and implementation of program to support them in reaching their potentials. WFP's Guidance for the Gender Marker has been in place since 2013, promoting the mainstreaming of gender in project proposals. Through an assessment of program documents provided by WFP China Office, gender-disaggregated data are available both at program objective and outcome reporting level, in terms of planned and actual numbers of children who received free school meals, and gender balance is achieved. Women are also the main beneficiary of income-increasing activity by providing home-grown food for preschools and employment in the kitchen of preschools. However, the program does not provide a detailed analysis of different nutrition needs and health issues facing different gender groups and their implications.

2.3.4. SUB-QUESTION 4: IS THE PROGRAM IMPLEMENTED AS EXPECTED?

128. The intervention had multiple components, namely: providing school meals, upgrading kitchen facilities, enhancing nutrition education of caregivers and preschool personnel and involving local poor smallholder farmers in the preschool food supply chain. To assess whether the program was implemented as expected, we reviewed a mix of annual progress reports, six-month progress reports, monthly monitoring

reports from China Development Research Foundation (CDRF) and smallholder farmer's focus group studies and key informants' interviews. In terms of the implementation fidelity, we found that the program was carried out as planned.

129.   <u>Providing school meals</u>. In 2018, the program provided 125,200 nutritious meals to 1,565 students[33]. The program fed more than the planned numbers of children in the following years. In 2019, the program provided meals to 2,552 children[34]. According to the program progress reports, there are 1,848 children who received school meals in spring semester and 1,792 in fall semester. This exceeded the planned 2,045 beneficiaries. In 2020, 2,586 children received school meals and 254,915 meals were provided[35]. During the spring semester in 2021, 151,066 meals (604,264 RMB) were provided, of which 592,028 RMB were for 1,896 children supported by WFP and 12,236 RMB for 41 children supported by the local government[36]. During the fall semester in 2021, 161,411 meals (645,644 RMB) were provided, of which 625,996 RMB were for 1,908 children supported by WFP and 19,648 RMB for 60 children supported by the local government[37].

130.   China Development Research Foundation (CDRF) provides monitoring support for the program using their digital technology and monitoring platform. Kitchen manager in each benefited preschool uploads key information of daily meals, such as menu and procurement data, to the platform. According to findings from CDRF, most of school lunches met the dietary intake standard, in terms of the intakes of cereals, meat, poultry, fish, soy products, and vegetables. The school meal also reached the standard nutritional intake of energy, protein, carbohydrates, iron, zinc, vitamin B1, vitamin B2, niacin, vitamin C and vitamin A. The only nutrient that did not meet the intake standard was calcium. In addition, excessive intake of sodium indicated excessive use of salt, therefore, the cooking method in the schools should be adjusted to reduce the use of salt.

131.   <u>Upgrading kitchen facilities</u>. Kitchen and canteen facilities were distributed and were used by preschools. In 2018, the program provided 722[38] pieces of kitchen and canteen equipment to treatment preschools, including refrigerated cabinets, refrigerators, disinfection cabinets, etc.

132.   <u>Enhancing nutrition education</u>. The program organized the first training in July 2018 with the support of the Chinese Center for Disease Control and Prevention (CDC). The nutrition knowledge messaging reached 43 teachers[39]. On top of the training sessions, 48 World Food Day posters and 142 Nutrition Knowledge Posters were distributed in the kindergartens connected to the project, potentially serving nutrition education purposes for children and parents. In 2019, 5 project management staff and 89 teachers attended the nutrition training workshops. 3,237 parents and grandparents attended 156 second-level community-based messaging sessions that were organized by the school teachers who had received the first level training[40]. In 2020, most nutrition education activities were moved online due to COVID-19 restrictions. 11 project staff, 250 teachers, and 3,639 caregivers participated in 231 nutrition education activities[41]. The program carried out 116 nutrition education activities in the spring semester of 2021 and reached 5 project

---

[33] WFP China Annual Country Report 2018

[34] WFP China Annual Country Report 2019

[35] Annual Progress Report 2020, Preschool Nutrition Improvement Programme

[36] 2021 Six-month Progress Report for Preschool Nutrition Improvement Programme (Xiangxi, Hunan Spring Semester)

[37] 2021 Six-month Progress Report for Preschool Nutrition Improvement Programme (Xiangxi, Hunan Fall Semester)

[38] 2018 Six-month Progress Report for Preschool Nutrition Improvement Programme (Xiangxi, Hunan Fall Semester)

[39] WFP China Annual Country Report 2018

[40] WFP China Annual Country Report 2019

[41] Annual Progress Report 2020, Preschool Nutrition Improvement Programme

staff, 120 teachers, 1,937 children and 1,937 caregivers[42]. In the fall semester, the program carried out 111 activities and reached 14 project staff, 126 teachers, 1,968 children and 1,863 caregivers[43].

133. <u>Involving smallholder farmers</u>. The program contracted with 76 smallholder farmers in 2018 and purchased an average of 1,552 RMB[44] worth of food from 50 farmers. The incomes of smallholder farmers contracted to supply the project have increased, with the project purchasing an average of 3,107 RMB worth of food from smallholder farmers in 2019 and an average of 4,252 RMB per household in 2020. Similar increases are witnessed in the number of farmers supplying the project (59 smallholder farmers in 2019 and 60 in 2020).

2.3.5. SUB-QUESTION 5: IS THE PROGRAM COST-EFFECTIVE?

134. In this section, we further conduct the cost-effectiveness analysis of this intervention following largely the methodologies discussed in McEwan (2012) and Dhaliwal et al. (2013). Cost-effectiveness analysis, in the simplest terms, calculates the ratio of the amount of cost required to achieve a given impact. It is defined as the ratio of the incremental cost per unit of incremental effect. For example, the cost of reducing the abnormal rate of hyperactivity by one percentage point. It should be noted that "incremental" means both the estimated impact and the incurred costs are relative to the comparison group. The lower the Cost/Effect ratio, the more cost effective the intervention. If the impact of the intervention on an outcome variable is not significantly different from zero, it implies the intervention is not cost-effective at all in terms of this outcome variable. In the rest of this section, we focus exclusively on those outcome variables on which the intervention is estimated to have a significant impact on average sample preschoolers or for certain subgroups of sample preschoolers.

135. We take a programmatic perspective in calculating the costs of the intervention by using the direct monetary program costs to the implementing organization. We are not taking a social perspective as we do not have data on the costs of public fund nor the costs incurred by households in responding to the interventions. In estimating the incremental costs from the programmatic perspective, we adopt the ingredients methods, which is a well-received method in the literature. Specifically, we calculate the programmatic costs during the intervention as the sum of six cost categories as follows: (a) administrative costs, (b) costs of school meal provision, (c) costs of milk provision, (d) kitchen facilities costs, (e) nutrition education costs, and (f) labor costs. For contents of cost items included in each cost category, please refer to Table 8 below. We measure the costs on a per capita (preschooler) basis by dividing the aggregate incremental costs by the number of beneficiary preschoolers in the treatment group.

Table 8 Contents of cost items included in each cost category

Administrative costs
(1) undertaking daily activities
includes: a) take pictures of the food every weekday
b) take pictures of how the beneficiary preschoolers enjoyed food at preschools every weekday
c) upload pictures to the Sunshine School Meals Platform
(2) material preparations and communications as required by the project
includes: a) materials to applying for the programme
b) materials about the executive progress of the programme
Costs of school meal provision
(1) school meal subsidy from the program during the intervention
(2) the incremental school meal fees from preschoolers' families in the treatment group relative to the comparison group during the intervention

Costs of milk provision
(1) the actual cost of the milk provided by this programme during the intervention (32 boxes/carton, 2.5 RMB/box, referred to the wholesale market price)
(2) delivery cost (including transportation and labor cost)
Kitchen facilities costs
∑(the total costs of one specific kitchen facility/the depreciable life of this kitchen facility a) * the intervention duration
Nutrition education costs
(1) organized the lectures to popularize nutritional knowledge
(2) organized and executed the daily nutrition knowledge promotion
Labor costs
extra labor cost in the intervention preschools compared with the compared preschools

Notes: a The kitchen facilities provided by this programme included tables, chairs, gas stoves, steam cabinets, freezers, refrigerators, sterilizers, scales, water dispensers, and sample retention cabinets. According to the national standard (i.e., Regulations of the People's Republic of China on the Implementation of Enterprise Income Tax), the depreciable life of the kitchen facility is set to 5 years (tables, chairs, steam cabinets, scales, water dispensers, and sample retention cabinets), 8 years (gas stoves and sterilizers), and 10 years (freezers and refrigerators), respectively.

136.   Recall that the impacts of this intervention on multiple outcome indicators, which measure the impacts on average treatment preschoolers or certain subgroups of treatment preschoolers (in the impact heterogeneity analyses) during the intervention. For examples, this intervention has been found to decrease the abnormal rate of hyperactivity among average treatment preschoolers by 4 percentage points during the intervention, or to decrease the anemia rate among female treatment preschoolers who were 3 years old at the time of baseline survey by 8 percentage points during the intervention.

137.   Table 9 presents the results regarding the incremental costs (Panel A), the incremental effect (Panel B)[45], and the Cost-Effect Ratio (Panel C) of the preschool nutrition pilot, respectively. As shown in Panel A, the incremental costs per treatment preschooler during the intervention is 4,221 yuan, which is consisted of 433 yuan, 1,287 yuan, 1,116 yuan, 23 yuan, 475 yuan, and 887 yuan for the administrative costs, school meal provision costs, milk provision costs, kitchen facilities costs, nutrition education costs, and labor costs, respectively.

138.   Results from cost-effectiveness analyses show that it costs 1,055 yuan to achieve one additional percentage point reduction in the abnormal rate of hyperactivity on average preschoolers (Row 18 in Panel C). As this is the only outcome variable on which the nutrition pilot is estimated to have significant impact when we look at the preschoolers as a whole, we then move onto those outcome variables with significant impacts for certain subgroups. To achieve one percentage point reduction in the prevalence of anemia, it costs 528 yuan per treatment schooler during the intervention for girls aged 3 years at the baseline whereas 1,055 yuan per treatment schooler during the intervention for boys aged 5 years old at the baseline (Rows 17, 19). In other words, the nutrition pilot is more cost-effective in reducing the anemia among 3-years old girls than among 5-years old boys (the latter costs almost twice as much as the former). To achieve one percentage point reduction in the prevalence of low verbal comprehension index (VCI), it costs 352 yuan per treatment schooler during the intervention for preschoolers from households which wealth index are at quintile 5 (Rows 15). To achieve one percentage point reduction in the prevalence of low working memory Index (WMI), it costs 422 yuan per treatment schooler during the intervention for preschoolers from households which wealth index are at quintile 4 (Rows 16). To achieve one percentage point reduction in the prevalence of wasting, it costs 1,055 yuan per treatment schooler during the intervention for preschoolers from poverty households (Rows 20). To achieve one percentage point reduction in the prevalence of stunting, it costs 2,111 yuan per treatment schooler during the intervention for boys (Rows 21).

---

[45] The impacts are taken directly from the impact evaluations in which effects are expressed as the average treatment effect on average treatment preschooler or on certain subgroups of treatment preschoolers.

Table 9 Cost effectiveness calculations

Panel A: incremental costs (RMB, per preschooler during the intervention)	
(1) Administrative costs	432.54
(2) School meal provision	1,287.19
(3) Milk provision	1,116.04
(4) Kitchen facilities	23.01
(5) Nutrition education	475.49
(6) Labor	886.78
Total costs	*4,221.06*

Panel B. impacts (percentage point, per preschooler during the intervention)	
Indicators, All sample	
(7) Abnormal rate of hyperactivity (%)	-4 (**)
(8) Other outcome indicators [a]	N.S. [b]
Indicators, Subgroup [c]	
(9) Low verbal comprehension index (VCI)rate (%), Children from households which wealth index are at quintile 5	-12 (**)
(10) Low working memory Index (WMI)rate (%), Children from households which wealth index are at quintile 4	-10 (*)
(11) Anemia rate (%), Girls-Age 3 at baseline	-8 (*)
(12) Anemia rate (%), Boys-Age 5 at baseline	-4 (*)
(13) Wasting rate (%), Children from poverty households	-4 (**)
(14) Stunting rate (%), Boys	-2 (*)

Panel C. Cost Effectiveness Ratio	
(15) Low verbal comprehension index rate (%), Children from households which wealth index is at quintile 5	351.75
(16) Low working memory Index rate (%), Children from households which wealth index is at quintile 4	422.11
(17) Anemia rate (%), Girls-Age 3 at baseline	527.63
(18) Abnormal rate of hyperactivity (%), all sample	1,055.26
(19) Anemia rate (%), Boys-Age 5 at baseline	1,055.26
(20) Wasting rate (%), Children from Poverty households	1,055.26
(21) Stunting rate (%), boys	2,110.53

Notes: [a] Other outcome indicators including the prevalence of anemia, under- and over-nutrition status, and social-emotional status, etc.
[b] N.S. means not significant.
[c] Since no significant effect was found for most outcome indicators in all sample, we further calculate the CEA for the subgroups with significant effects including the prevalence of anemia, stunting, wasting, low WMI and low VCI.

139.   When we look at the composition of the incremental costs, our data show that administrative costs and labor costs together account for nearly 1/3 of the aggregate incremental costs. It might be interesting to compare the cost composition of similar WFP programs in similar contexts to see if there is still room make the intervention more cost-effective from the administration and labor perspectives.

140.   The results also revealed that among those outcome variables with significant impacts for certain subgroups of treatment preschoolers, their cost-effectiveness ratios vary significantly. In other words, even with the same intervention, the costs vary a lot to achieve a given impact on certain outcome variables for certain subgroups of beneficiaries. For example, results from cost-effectiveness analyses show that the nutrition pilot is twice as cost-effective in reducing the anemia among 3-years old as among 5-years old boys.

These research findings have important implications for the targeting of similar nutrition interventions in the future.

3. Conclusions and Recommendations

141.   Based on the findings presented in the previous section, an overall assessment that responds to the evaluation questions is provided below, followed by recommendations for future research and how WFP can take action to improve outcomes based on our findings.

3.1. CONCLUSIONS

142.   We find that the project's direct purchase from smallholder farmers contributed to local smallholders' vegetable production diversity, dietary diversity, income, nutrition knowledge and market engagement. Outcomes related to child nutrition and development portray a more mixed picture. Measures of child nutrition, cognition and social emotions showed improvements from the baseline survey to the endline survey. However, while the improvement on some of these indicators was greater in treatment groups than comparison groups, this difference was not statistically significant. A strict interpretation of these findings is that the project had no impact on nutrition and development of preschool children. Note that the Project was evaluated against a backdrop of China's national momentum towards reducing child malnutrition and widespread improvements in key maternal and child health indicators. This was also one of the explanations that the project did not show impacts as child nutrition in both treatment and comparison groups are improving, possibly thanks to the national efforts. The conclusions below are organised around the broad headings within the ToR.

143.   This study focused on Longshan county and Yongshun county of Xiangxi Tujia and Miao Autonomous Prefecture, Hunan province of China. Both of them were nationally designated poverty counties in 2018. As for potential generalisability of our study, we expect that our evaluation findings would be most relevant for children in poor rural areas in South Central China. However, it might not reflect the situation in areas with higher level of development and other regions. Whether the same results would be obtained if the programme were extended to more severely disadvantaged areas is also not certain. Future research endeavours should consider expanding the scope to include more regions of China to enhance national representativeness.

Evaluation Question 1. What is the impact of nutrition program on nutrition and development of preschool children aged 3 and 5?

144.   **Nutrition:** Overall, we found no impact of the project on child nutrition outcomes. From the baseline to the endline, the prevalence of anemia declined significantly. However, this decrease was similar in treatment and comparison groups. The project therefore had no impact on reducing anemia; The project also had no impact on anthropometric outcomes (stunting, wasting, and underweight) although few children were malnourished in the population studied during endline. We found some evidence of impacts of the intervention on children's nutritional status of sub-groups. For example, the program had a significant impact on the decline of anemia rates in 3-year-old girls and 5-year-old boys; and on the decline of stunting prevalence in boys. However, these effects were heterogeneous, depending on age, gender, and poverty status. The few significant impacts did not follow any clear pattern and we could not conclude the program led to systematic improvements across all nutritional outcomes for any of these subgroups. At the other end of the spectrum, it appears that the project did have a nontrivial impact on the likelihood of child being overweight.

145.   **Cognition:** The analysis uncovered that the project had no impact on children's cognitive ability. However, results from heterogeneity analyses showed that the project improved the cognition of children from economically better-off families.

146.   **Social Emotions**: Similarly, we found no impact of the project on children's social emotion status overall. However, we found it helped reduce the abnormal rates of hyperactivity of children on average. The project also turned out to be effective in reducing the abnormal rate of total difficulties for some sub-groups of children. Specifically, it significantly reduced the abnormal rates of hyperactivity and total difficulties in the younger children; and helped reduce abnormal rates of emotional symptoms and hyperactivity for children from poor families.

147.   We examined some key factors that constrain the positive school feeding effect on nutrition and development of preschool children, including nutrition knowledge score, feeding practice and interactive parenting. Results show that the level of nutrition knowledge of primary caregivers is inadequate. Dietary quality at home and in preschools has improved in both groups similarly. In other words, the school feeding program had little impact on children's dietary quality. Interactive parenting quality was also similar between groups. Part of the explanation likely lies in the timing and focus of the attention on activities devoted to nutrition relative to behaviour change messaging for parenting practices.

148.   The lack of difference between treatment and control groups could also arise because of spillovers in knowledge and practices. If the project improved nutrition practices in treatment schools that were observed and copied by comparison schools, children from the comparison group could see improvements as well.

Evaluation question 2. What is the impact of nutrition program on local smallholders?

149.   We found that direct purchase from smallholder farmers contributed to local smallholders' vegetable production diversity, income, nutrition knowledge and market engagement. In addition, based on analysis from the 106 households (endline survey), we found that smallholders in the treatment group had a higher dietary diversity score (DDS) than those in the comparison group. This program also has the potential to assist in building a local food system that is resilient, sustainable and inclusive.

150.   We also identified several constraints to be overcome for sustaining or scaling up the current pilot. These challenges include the unstable supply under the farm to preschool model, the insufficient supply-side intervention, the transportation constraints and the expected decrease of demand from the local preschools. It is worth mentioning that the targeted beneficiaries of the smallholders varied in endline survey from the baseline survey. According to the initial selection criteria of smallholders in 2018, beneficiary households should meet the following qualifications: 1) registered poor households; 2) capacity (e.g. land, labor) to engage in agricultural production; 3) a child attending the preschool with which they are about to sign the contract; and 4) signed agreement on a voluntary basis. The aforementioned criteria imply that the treatment group smallholders included in the baseline survey have to withdraw from the program when their children or grandchildren graduated from preschools. This led to high turnover rates of the beneficiaries. As a result, it is important to note that the results presented in those sections rely on very small sample sizes which reduce the accuracy of the indicators measured and the magnitude and direction of observed changes over time needs to be interpreted with caution.

Evaluation question 3. How appropriate is the program implementation?

151.   The program design and implementation are consistent with policies, strategies, and guidance on nutrition, child development, and poverty reduction in relevant national, UN and WFP level.

152.   When we look at the composition of the incremental costs, our data show that administrative costs and labor costs together account for nearly 1/3 of the aggregate incremental costs. It is worthwhile to compare the cost composition of similar WFP programs in similar contexts to see if there are rooms to make the intervention more cost-effective from the administration and labor perspectives.

153.   Cost-effectiveness ratios varied under the different intervention effects and costs input in different subgroups. For example, results from cost-effectiveness analyses show that the nutrition pilot is twice as cost-effective in reducing the anemia among 3-years old as among 5-years old boys.

3.2. RECOMMENDATIONS

154.   Based on the findings and conclusions of this evaluation, the recommendations of the evaluation team are outlined below. The various recommendations are inter-related and are best considered as a comprehensive package.

155.   **Recommendation 1: Building on its experience and reputation, the WFP should extend the provision of interventions in the study sites.** Despite improvements from baseline and endline, there is still a gap in terms of nutritional outcomes between children in our study and the national average. Social emotional status of these children was also worse than their peers according to the results from other Chinese context researches, indicating potential for further improvements. In the future, the focus of WFP

should also include the sustainability of project, e.g., how to increase the participation of local government without WFP support.

156. **Recommendation 2: The WFP should consider targeting at vulnerable sub-groups to optimize the impact of the project.** Our findings on heterogeneous project effects across sub-groups pointed out the need to look into differences between children with diversified demographic and socioeconomic statuses. Particularly, attention needs to be paid to children from disadvantaged backgrounds (left-behind children, children from registered poor households, etc.) during the program design and targeting. In high- and middle-income countries free school meals are generally targeted to individual children on the basis of vulnerability and well-being indicators. It is recommended for WFP to consider targeting to ensure that project generates a maximum impact to the intended beneficiaries, especially given a finite budget.

157. **Recommendation 3: The WFP should pay particular attention to the rise of child overweight and obesity**. Under the large and rapid increase in the overweight and obesity rate, it's recommended that the WFP should invest in nutrition standards for school meals and ensure healthier meals in the field. Considerations include calorie intakes, nutritional values, requiring more fruits, vegetables and whole grains, and less sodium and saturated fats.

158. **Recommendation 4: The WFP needs to take into consideration nutrition knowledge of caregivers to improve children's diet at home in addition to the nutritional value of the food choices offered in schools.** The gap between perceived nutrition knowledge and actual increase of nutrition knowledge need to be taken into account in the social behavior change communication (SBCC) activity in this project. The fact that caregivers had an improved yet only minor shift in knowledge score points to the fact that there is a need to continue and improve trainings on nutrition knowledge. While additional efforts should be made to increase parents' participation in the meetings, the inflexible nature of a set meeting time will realistically limit some parents' abilities to attend due to conflicting responsibilities, thus it is also recommended that future programming could identify additional opportunities to reach parents and caregivers with literacy supportive messages. The trainings should pay explicit attention to increase female caregivers' involvement given our findings on the disparity of mothers'/female caregivers' nutrition knowledge.

159. **Recommendation 5: The WFP should support further research to understand reasons for the decline in child malnutrition.** The decline in the prevalence of child malnutrition is remarkable. In this evaluation study, the prevalence of anemia has declined from 32% to 6%. Our findings show that the beneficial impacts attributable to school feeding are limited if we attempt to extract school feeding from the larger context of China's national momentum towards reducing child malnutrition and widespread improvements in key maternal and child health indicators. Therefore, it would be very useful to explore whether the decline was due to a national policy, macroeconomic improvements, a targeted policy or project, or due to other factors. Related, it is important to understand how WFP project is aligned and interacted with these factors.

160. **Recommendation 6: The WFP should strengthen supply-side interventions to support local smallholders.** First, continuous interventions are needed to stimulate smallholder's participation both in the school feeding and other markets beyond the school feeding project. Second, given findings on the decreasing demand of the preschools, the current pilot project could be extended to a broader public procurement, such as public demand from local governments and primary schools. Third, providing more school-related job vacancies such as preschool caterers and cooks may also be an avenue to increase local smallholders' income.

161. **Recommendation 7: The WFP should invest in rigorous evaluation at both country office and field level.** Rigorous evaluation is important to demonstrate the outcomes and impact of projects. It is recommended that the WFP country office should invest in evaluation exercises to generate evidence and inform future project design. At the field level, child-level data exists in large hardcopy ledgers and have not been collated for evaluation purposes, such as student's attendance, therefore impossible to track multiple entries for the same child. The data in question are collected as a necessary condition of the project taking place, but currently cannot be used in the evaluation process. Efforts could be made to digitize this data collection and to track children between periods of receipt of interventions.

Annexes

Annex 1. Maps (in Chinese)

Hunan Province in China

中华人民共和国

审图号：GS（2017）2166号　　　比例尺 1：29 000 000
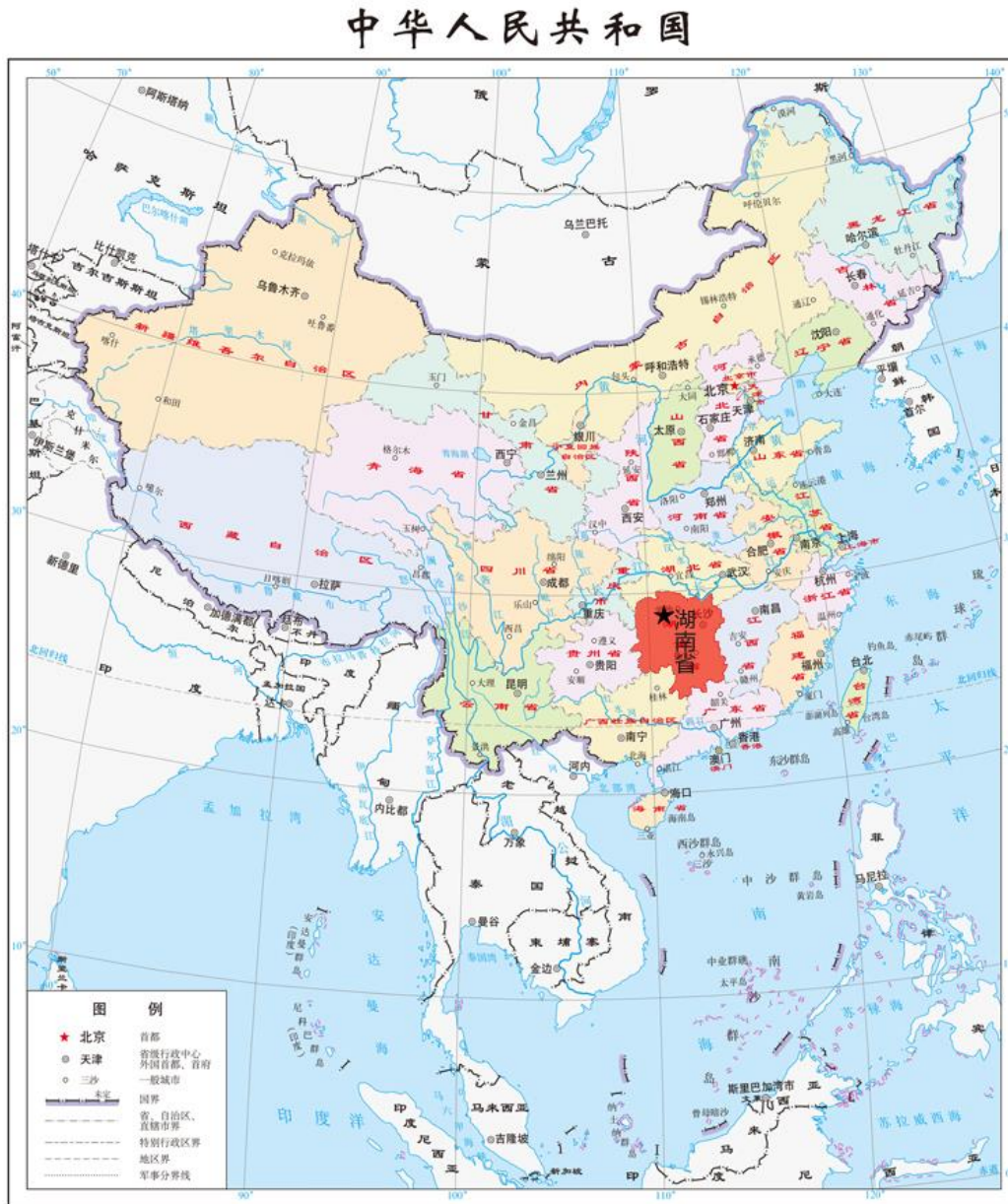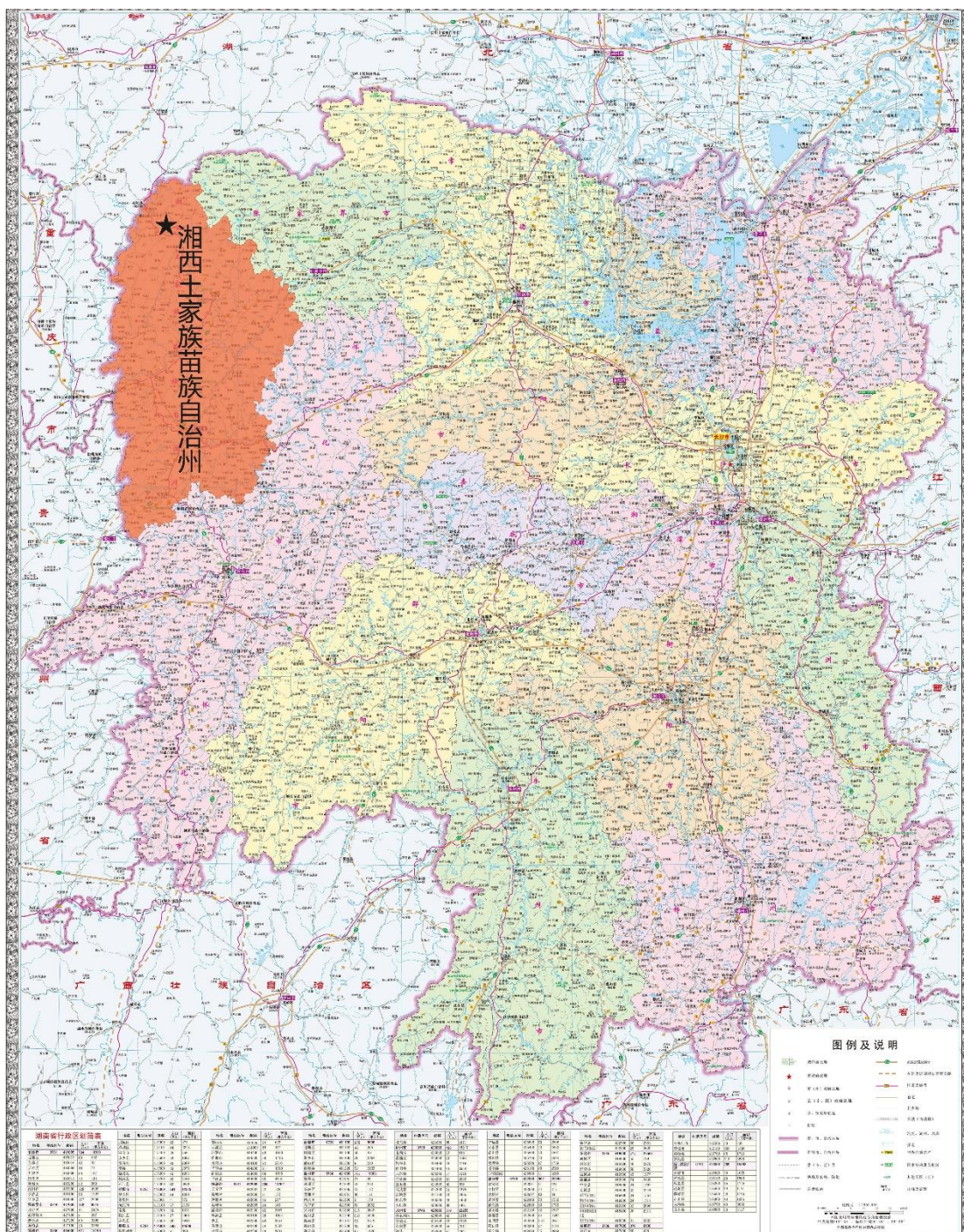
Selected Longshan and Yongshun counties in Xiangxi Prefecture

Selected townships in Longshan county

Selected townships in Yongshun County

Annex 2. Timeline

Responsible Stakeholders	Phases, Key Activities, and Deliverables	Duration/Dates
Inception phase		
All the internal and external stakeholders	Conduct scoping and pre-test surveys	May 31-June 1; June 8-13; July 16-17, 2018
	Organize an inception meeting with key stakeholders	Mid-August, 2018
	Finalize evaluation design	Mid-August, 2018
	Deliverable: Inception report	07/31/2018
Baseline Phase		
Internal WFP stakeholders Beneficiaries Local governments UN country team Private sector	Conduct the baseline survey	September 1-15, 2018
	Analyze data and draft baseline report	September, 2018-February, 2019
	Arrange a debriefing session for the baseline findings	Mid-January, 2019
	Deliverable: Baseline report	02/28/2019
Endline Phase		
Internal WFP stakeholders Beneficiaries Local governments UN country team Private sector	Conduct the endline survey	May 14-29, 2021 (delayed)
	Analyze data and draft endline report	June, 2021- October, 2021
	Arrange a debriefing session for the endline findings	October, 2021
	Deliverable: Endline report	11/30/2021
	Organize a dissemination meeting with key stakeholders	To be discussed
	Deliverable: 1) Final evaluation report; 2) Policy briefs	To be discussed

Annex 3. Theory of Change

1. The Theory of Change (ToC) is mapped out to show the activity-output-outcome-impact causal chain through which the program objective of improving child nutrition and the well-being of smallholders can be achieved. Arrows are intended as an approximate representation of causality, but this is only schematic.

2. We also specify key assumptions that underlie the casual chain. All these assumptions are represented by the small numbered boxes in the ToC diagram below. For reasons of graphical simplicity, they could not all be positioned optimally, and it could well be suggested that some of them could more meaningfully be placed elsewhere. In several cases, one assumption is shown at several places, indicating the multiple points in the ToC to which it is relevant.

3. The evaluation questions clarified in the matrix at Annex D are developed based on the ToC with considerations that 1) activity, output, outcome, and impact illustrated in the ToC are different levels of changes throughout the program, displaying a reasonable set of early and intermediate steps toward the long-term goal. All the levels of changes, which could provide insights in why long-term goals are realized or not, deserve further examination; and 2) assumptions identified in the ToC are underlying issues that could either confound or facilitate the progress of the program. The validity of those assumptions need test to see whether the efficiency, effectiveness, and impact of the program could be achieved.

Activities phase-Key assumptions:

1) Funds are available on time.
2) Program staff are appropriately qualified and motivated.
3) Program policies, strategies and procedures are not only formulated but also meaningfully implemented.
4) Information about preschools, caregivers, and smallholders are correct and are fully understood by the program staff.
5) All the caregivers, teachers, administrators, and kitchen staff attend the nutrition training/lecture as required.
6) Smallholders fully understand and strictly follow the procurement agreement. Necessary mechanisms are set up to ensure the quantity, quality, and safety of their products.

Outputs phase-Key assumptions:

7) Foodstuffs are available in sufficient quantity and quality at right time, and are properly prepared and distributed to children.
8) Kitchen and canteen facilities with sufficient quantity and quality are distributed on time and are used by preschools.
9) Education facilities with sufficient quantity and quality are distributed on time and are used by preschools.
10) Teachers, administrators, and kitchen staff value the nutritious meal and make sure that children eat the meals properly.
11) Caregivers, teachers, administrators, and kitchen staff internalize what they learn from the training/lecture, value and are willing to practice and spread the new knowledge.
12) Personnel changes of preschools will not be at a level that diminishes the effectiveness of capacity development.
13) Children understand what they learn from teachers and caregivers, value and are willing to practice the good eating habits.
14) Caregivers view the school meal as a complement rather than substitution of family feeding. That said, the quantity and quality of food which preschoolers eat at home will not be reduced.

Outcomes and impacts phase-Key assumptions:

15) The causal links between preschool meals, enhanced children nutrition, and improved cognition and emotion are correct.
16) The program receives sufficient attention and support, and is implemented thoroughly as designed.
17) External factors affecting beneficiary livelihoods, such as economic performance at the macro level, remain stable throughout the program.

Inputs and

Output

Outcomes

Impact

Provide smallholders with inputs such as seeds, fertilizer, and pesticide

Increased availability of key inputs such as seeds, fertilizer and pesticide at smallholder level

Improved smallholders' agricultural technical capacity

Organize agricultural technical training for smallholders

Enhanced production efficiency of smallholders

Increased smallholders' production

Agricultural growth

Sign procurement agreements with smallholders

Increased demand for food associated with preschool feeding program, which is highly structured and predictable.

Increased smallholders' engagement with food markets

Increased smallholders' income

Poverty reduction in rural

Increased dietary diversity

Improved nutritional status of smallholders

Assumptions

- Agricultural growth is essential for board-based poverty reduction in the rural area.
- Agricultural growth will come about through increased engagement with input and output markets, which will stimulate technical change and result in productivity enhancement.
- The greatest poverty reduction impact of this "market engagement-agricultural growth-poverty reduction" linkage will come via a focus on smallholders.
- The major constraint to market engagement by smallholders is that input and output markets are thin, seasonal and poorly governed.
- The market constraint can be addressed by using the public demand for food associated with school feeding program to drive a demand-assisted agricultural growth strategy.
- The demand for this school feeding program is highly structured and predictable and thus amenable to a structuring process that will reduce the uncertainty and risk associated with smallholders' engagement with food markets.
- The provision of complimentary services which include training and access to inputs and technology can be linked to the process of structured demand.
- Thus, the use of structured demand to supply school feeding program is an attractive avenue through which to kick start a process of agricultural and livelihood transformation.

Annex 4. Evaluation Matrix

Evaluation question 1 What is the impact of nutrition program on nutrition and development of preschool children aged 3-5?				Impact, Relevance, Effectiveness
Subquestions	**Indicators**	**Data collection methods**	**Sources of data/information**	**Data analysis methods/triangulation**
To what extent does nutrition program improve children's nutrition status?	Weight, height, and haemoglobin level (benchmark: national averages)	Questionnaire survey to record the birth weight and body length of children; Physical examination through which professional nurses will collect the information; Secondary data collection	Primary data from survey; Government statistics (secondary data)	Descriptive analysis; Difference in differences; Cross-check findings from field survey with recorded data gathered by local health department and informants from government
To what extent does nutrition improve children's cognitive level?	Score of WPPSI-IV and WISC-IV Test (benchmark: percentile rank will be calculated for each index score)	WPPSI-IV and WISC-IV Test through which trained enumerators will test the cognitive level of children	Primary data from survey	Descriptive analysis; Difference in differences

To what extent does nutrition improve children's socio-emotional status?	Score of Ages & Stages Questionnaires; Score and abnormal rate of Strengths and Difficulties Questionnaires	Questionnaire survey in which there are sections about socio-emotional development	Primary data from survey	Descriptive analysis; Difference in differences
Sub-question 4. To what extent did the nutrition program improve the nutrition knowledge of preschool personnel and caregivers?	Analysis of the nutrition knowledge section in the field survey questionnaires	Questionnaire survey with specific questions to test the nutrition knowledge of participants; Key informants interviews	Primary data from survey; Informants from local governments, and academic institutions	Descriptive analysis; Difference in differences
Sub-question 5. To what extent did the nutrition program improve the food consumption and nutrient intake of children?	Analysis of the food expenditure and diet record section in the field survey questionnaires	Questionnaire survey with specific questions for food expenditure and diet record	Primary data from survey	Descriptive analysis; Difference in differences
Sub-question 6. To what extent did the nutrition program improve the interactive parenting of caregivers?	Analysis of interactive parenting module in the field survey questionnaires	Questionnaire survey with specific questions to test the quality of interactive parenting	Primary data from survey	Descriptive analysis

Evaluation question 2. What is the impact of nutrition program on local smallholders?				Impact, Relevance, Effectiveness
Subquestions	**Indicators**	**Data collection methods**	**Sources of data/information**	**Data analysis methods/triangulation**
To what extent does the nutrition program boost agricultural production of smallholders?	Land cultivated; Agricultural technology training	Questionnaire survey with specific questions for agricultural production	Primary data from survey	Descriptive analysis; Difference in differences type of comparison
What is the impact of nutrition program on production diversification of smallholders?	Production diversity score; Crop diversity score; Livestock diversity score	Questionnaire survey with specific questions for agricultural production	Primary data from survey	Descriptive analysis
To what extent does the nutrition program contribute to local smallholders' income?	Income per capita	Questionnaire survey with specific questions for income composition	Primary data from survey Focus group discussion	Descriptive analysis; Difference in differences type of comparison; case study
What is the impact of the nutrition program on dietary diversity of smallholders?	Dietary diversity score	Questionnaire survey with specific questions for food expenditure and diet record	Primary data from survey Focus group discussion	Descriptive analysis; case study

Evaluation question 3. How appropriate is the program implementation?				Efficiency, Sustainability, Effectiveness
Subquestions	**Indicators**	**Data collection methods**	**Sources of data/information**	**Data analysis methods/triangulation**
To what extent are the program's designed objectives, targeting strategies, and activities consistent with relevant national policies, strategies, and guidance on nutrition, child development, and poverty reduction?	Alignment of the program's designed objectives, targeting strategies, and activities with those stated in relevant national policies, strategies, and guidance	Document review; Key informants interview	Documentation (secondary data) of program design and implementation from the WFP, as well as documentation of relevant policies, strategies, and guidance from national, provincial, and local governments; Key informants from the WFP, various levels of governments, private sector, and academic institution	Analyze program reporting and government documentation to check the coherence, comparing with views of informants
To what extent are the program's designed objectives, targeting strategies, and activities consistent with UN and WFP policies, strategies, and guidances on nutrition, child development, school feeding, and poverty reduction?	Alignment of the program's designed objectives, targeting strategies, and activities with those stated in UN and WFP policies, strategies, and guidance	Document review; Key informants interview	Documentation (secondary data) of program design and implementation from the WFP, as well as documentation of relevant policies, strategies, and guidance from UN system and the WFP; Key informants from the WFP and other UN agencies	Analyze program reporting and UN documentation to check the coherence, comparing with views of informants

Is the program designed and implemented based on a sound gender analysis that is in line with national, UN, and WFP gender policies, considering the distinct needs and participation of boys/males and girls/females?	Alignment of the program's design with those stated in national, UN, and WFP gender policies	Document review; Key informants interview; Questionnaire survey	Documentation (secondary data) of program design and implementation from the WFP, as well as documentation of gender policies from various levels of governments, UN system and the WFP; Key informants from governments, the WFP, other UN agencies, private sector, and academic institutions; Primary data from field survey	Analyze program reporting, as well as government and UN documentation to check the coherence, comparing with views of informants and findings in the baseline survey
Is the program cost-effective?	Program's cost-effectiveness ratio	Questionnaire survey; physical examination; Cognitive test; Document review; Key informant interviews	Primary data from questionnaire surveys, physical examination and cognitive test; Documentation (secondary data) from WFP, local government, private sector, and academic institution; Key informants from the WFP, local government, private sector, and academic institution	Summarize expenditure types; Calculate cost-effectiveness ratio

Annex 5. Data Collection Tools

1. First and second pre-test surveys during July-August, 2018: The evaluation team conducted extensive pretesting of the surveys to make sure that they ask the questions in a clear, straightforward way, and include both easy and difficult questions. This helped ensure that the surveys capture an accurate distribution of caregiver understanding, rather than underestimating understanding by asking confusing or misleading questions or overestimating it by asking easy questions.

2. Baseline survey in September, 2018

   1) Child survey: 1,334 kids. Major modules of the survey include the following

      - Physical examination by professional nurses: height, weight, and hemoglobin test

      - Cognition test by trained enumerators: Using Wechsler Preschool and Primary Scale of Intelligence-IV (WPPSI-IV)

   2) Child's primary caregiver survey: caregivers of the 1,3344 kids. Major modules of the survey include the following

      - Basic information of the child: date of birth, gender, ethnicity, health, time spending with parents, etc.

      - Child's living environment: parents' marital status, indoor smoking at home, etc.

      - Basic characteristics of each household member: age, gender, ethnicity, education, occupation, migration, health, marriage, nutrition training or education, etc.

      - Basic information of the child's preschool education: enrolment date, means and time of transportation to preschool, tuition and fees, caregiver's overall evaluation of the preschool, etc.

      - The family's investment on childrearing and the caregiver's expectation of the child's educational achievement

      - Food consumption of the child at home

      - Caregiver's knowledge on child's oral hygiene

      - Socio-economic status of the household: position of durable assets, etc.

      - Nutrition knowledge of caregivers

      - Parenting practices of caregivers

      - Socio-emotional behaviour of the child: based on internationally recognized Ages & Stages Questionnaires, completed by the caregivers

      - Physical examination by professional nurses: height, weight, and hemoglobin test

   3) Teacher survey: teachers of the 26 preschools. Major modules of the survey include the following

      - Basic characteristics of the teacher: age, gender, ethnicity, marriage, health, education, teaching experiences, rank, salary, etc.

      - Nutrition training the teacher received

      - Class characteristics: class size, composition in terms of age, gender, ethnicity, left-behind status, poverty status, disability, communication with caregivers through Wechat, ect.

      - Class facilities and curriculum

      - Teaching practices and class management

      - Food consumption, performance, and attendance of sample children in preschool

      - Nutrition knowledge of the teacher

- Childrearing practices of the teacher
- Physical examination by professional nurses: height, weight, and hemoglobin test

4) Principal survey: principals of the 26 preschools. Major modules of the survey include the following

- Basic characteristics of the preschool: the date of founding, location, ownership, subsidy from the government, facilities, environment of the preschool, etc.
- Size of the preschool: number of classes, children's enrolment situation in terms of grade, age, gender, ethnicity, left-behind status, poverty status, and disability, etc.
- Tuition, fees, and subsidy
- Preschool facilities and curriculum
- Teachers and their composition in terms of age, gender, ethnicity, rank and marriage
- Basic characteristics of the principal: age, gender, ethnicity, education, teaching experiences, rank, marriage, health, nutrition training, etc.
- Nutrition knowledge of the principal
- Interactions between preschools and caregivers
- Physical examination by professional nurses: height, weight, and hemoglobin test

5) Kitchen manager survey: kitchen manager of the 26 preschools. Major modules of the survey include the following

- Basic information on preschool kitchen: facilities, subsidies, number of staff, nutritional training for staff, etc.
- Information on preschool meals: number of meals provided, number of children and staff eating at preschool, types of foodstuffs provided
- Procurement of the kitchen
- Basic characteristics of the kitchen manager: age, gender, ethnicity, education, teaching experiences, rank, marriage, health, nutrition training, etc.
- Nutrition knowledge of the manager
- Physical examination by professional nurses: height, weight, and hemoglobin test

6) Smallholder survey: 76 selected smallholders and other 76 in the comparison group. Major modules of the survey include the following

- Basic characteristics of the household: age, gender, education, occupation, and health of each family member; previous technical training on agricultural production; whether the family has preschool kids, etc.
- House and fixed assets of the household
- Agricultural production: land; vegetable and fruit production; poultry, fish and livestock production
- Detailed household income and expenditure
- Household food consumption yesterday and in the past week
- Information on agricultural production contract, market, logistic, cooperatives, and participant's nutrition knowledge

3. **Third pre-test survey** during April, 2021: The evaluation team tested all the questionnaires to make sure that each question can be clearly understood by the interviewee and the accurate information is fed back, particularly for the cost questionnaires which were first used in the field survey. Anothor mission

was to ensure the validity of electronic questionnaires which were designed for the Computer-Assisted Personal Interviews (CAPI) methods. The computerized devices applied in this method were also tested to assess the stability, debug and make technical protocols. This survey helped capture the progress of the program as well as the potential challenges for the following evaluation process.

1. **Where-about survey** in April, 2021: To capture the where-about information of each sample child, the survey team visited each of the 26 sample preschools immediately before the endline survey. During this survey, the potential attrition rate and its influence to the evaluation process was estimated. Based on the findings, the evaluation team make optimal plans for endline survey to minimize the negative impacts of survey delay.

2. **Endline survey** in May, 2021: During the endline survey, the evaluation team returned to the sample preschools to survey the same 1,334 children as well as their caregivers, teachers, administers. These children who graduated from preschool or transferred to other preschools were also traced as far as possible. Meanwhile, the evaluation team went back to the 152 smallholders for information on smallholders' participation.

   1) The contents of the survey almost the same as the baseline survey in 2018 so that the evaluation team were able to have a panel dataset to better evaluate the impact of the project.

   2) Cost questionnaire was used to guide the interview with key informants from stakeholders including the WFP, provincial and county governments, preschools/preschools, benefited households, private sector, and nongovernmental organizations, to collect their cost data or estimate relevant costs.

      • Collecting WFP cost data: WFP will keep records on the costs of its program, including 1) direct program costs: free meals for children, facilities/nutrition trainings for preschools, agricultural inputs/technical trainings for smallholders; 2) direct support costs: facilities for program management, domestic and international travels/field missions/internal meetings and trainings regarding the program, salary for WFP staff; and 3) other indirect support costs: WFP overhead, etc.

      • Collecting government cost data: Provincial and especially county governments will directly provide financial support to free meals and agricultural inputs/technical trainings for smallholders. This part of costs will be collected. To obtain data on other costs generated by facilities, travels/field missions/internal meetings and trainings regarding the program, salary for government staff, and other costs contributed by governments, the evaluation team will interview government officials responsible for the nutrition program.

      • Calculating preschool-level cost for both treatment and comparison groups: Besides direct costs on the WFP and government sides, providing children with free meals will also involve costs from the benefited preschool itself, such as inputs in meals, kitchen, and facilities as complements of the program. Preschool principal, teachers, and kitchen staff will be asked to estimate and monetize the increased amount of time they contributed to preschool feeding.

      • Estimating household-level cost for both treatment and comparison groups: Costs generated through family feeding and childrearing practices will be estimated in the field surveys.

      • Collecting cost data from private sectors: Private sector representatives involving in the nutrition program will be interviewed to collect their costs on free milk and nutrition education activities.

      • Collecting cost data from nongovernmental organizations: Representatives from academic institutions and NGOs involving in the nutrition program were interviewed to collect their costs on nutrition training and the management of data sharing platform.

Annex 6. Fieldwork Agenda

Days/dates	Field Missions	Team members	Locations	Activities	Stakeholders
May 31-June 1 2018	Scoping mission	Dr. Kevin Chen Ms. Zimeiyi Wang	Yongshun county	Collect relevant information; Identify potential opportunities and challenges facing the surveys; Develop a working relationship with local program offices	China office of WFP; beneficiaries; government representatives at all levels; private sector representatives
June 8-13 2018	Pretest survey (first)	Dr. Chengfang Liu Ms. Zimeiyi Wang Ms. Yanying Yu	Baojing Longshan Yongshun	Collect relevant information; Pretest draft questionnaires; Fine-tune the sampling strategy; Identify potential opportunities and challenges facing the surveys; Train the enumerators team leaders; Develop a working relationship with local program offices	China office of WFP; beneficiaries; government representatives at provincial and county level
July 16-17 2018	Pretest survey (second); Nutrition training observation	Ms. Zimeiyi Wang	Jishou city	Pretest draft questionnaires; Identify potential opportunities and challenges facing the surveys; Take observation notes on the training delivery; Test trainees on their nutrition knowledge before and after the training	China office of WFP; beneficiaries; government representatives at provincial and county level; academic institutions

Days/dates	Field Missions	Team members	Locations	Activities	Stakeholders
September 1-15 2018	Baseline survey	Dr. Kevin Chen Dr. Chengfang Liu Ms. Zimeiyi Wang Ms. Yanying Yu	Longshan Yongshun	Conduct questionnaire survey and case study	Internal WFP stakeholders; beneficiaries; local governments; UN country team; private sector
April 19-22, 2021	Pretest survey (Third);	Dr. Kevin Chen Dr. Chengfang Liu Ms. Zimeiyi Wang Ms. Yanying Yu	Longshan Yongshun	Pretest draft questionnaires and interview outlines; Identify potential opportunities and challenges facing the surveys; Test efficiency and applicability of the electronic questionnaires.	Internal WFP stakeholders; beneficiaries; local governments; private sector
April 22-24, 2021	Whereabout survey	Ms. Yanying Yu	Longshan Yongshun	Collect accurate whereabout information about all sample children; Estimate the potential attrition rate and its influence to the evaluation process; Prepare optimal plans for endline survey to minimize the negative impacts of survey delay.	Internal WFP stakeholders; beneficiaries; local governments; private sector
May 14-29, 2021	Endline survey	Dr. Kevin Chen Dr. Chengfang Liu Ms. Yanying Yu	Longshan Yongshun	Conduct questionnaire survey and case study	Internal WFP stakeholders; beneficiaries; local governments; UN country team; private sector

Annex 7. Methodology

Annexes 7.1. Identification Strategy for Children's Nutrition and Development (FOR EQ1)

We estimated both intention to treat effects (ITT) and average treatment effects (ATE) to measure the impact of project on children's nutrition and development. ITT results reported cover both the students who participated in the project as well as those who were assigned to treatment but either went to alternative preschools or else dropped out of schooling.

For continuous variables, we used a Difference-in-Differences (DID) model in the following form to estimate the ITT effect:

$$\Delta Y_i = \alpha + \beta_1\, T + \beta_2\, Y_{i1} + \beta_3\, C_i + \varepsilon_i$$

where ΔY_i is the outcome difference between two waves, Y_{i1} is the outcome variable at baseline and T is a dummy variable for the treatment assignment. β_2 is the DID estimator which we focus on.

For dichotomous variables, we followed like Liu et al., 2017, Mbuya et al., 2019, and Olney et al., 2015 and used endline outcomes as the dependent variables in the following form to estimate the ITT effect:

$$Y_{i2} = \alpha + \beta_1\, T + \beta_2\, Y_{i1} + \beta_3\, C_i + \varepsilon_i$$

where Y_{i1} is the outcome variable at baseline, Y_{i2} is the outcome variable at endline.

We estimated ATE which reflect the impact of the intervention on those children who actually took up treatment, by using the treatment assignment as an instrumental variable for actual treatment status. We performed OLS for continuous outcome variables. For dichotomous quantifiers, we performed probit models. We also include subgroup analysis by age, gender, and household poverty.

Annexes 7.2. Methodology for smallholder module (FOR EQ2)

Sampling

In the baseline survey, 146 households were interviewed. Of which, 72 households were from the treatment group whereas the rest 74 from the comparison group. In the survey, we collected detailed information on their household characteristics, agricultural production, income, consumption, etc. The baseline data showed that farmer households in the treatment and comparison groups are balanced in terms of their characteristics.

In the endline survey, 106 smallholders were interviewed. Of which, 58 belonged to the treatment group whereas the rest 48 the comparison group (Table 7.1). The 58 treatment group smallholders were those who were supplying food to preschools at the time interviewed. It is interesting to note that 12 of the 58 smallholders in the treatment group were those supplying to preschools at the time when the baseline survey was conducted.

The comparison group are comprised of those smallholders in the baseline survey whom we managed to track during the endline survey (27 households). In addition, 21 households were newly included in the endline survey. They are from caregivers of preschools in the comparison group which did not receive any intervention from this nutrition program.

Table A7.1 Composition of sample smallholders in the endline survey (n)

Participated in baseline or not?	Treatment Group	Comparison group	Sub-total
Yes	12+2 (transferred from baseline comparison group)	25+2 (transferred from baseline treatment group)	41
No	44	21	65
Sub-total	58	48	106

It is worth mentioning that the targeted beneficiaries of the smallholders varied in endline survey from the baseline survey. According to the initial selection criteria of smallholders in 2018, beneficiary households should meet the following qualifications: 1) registered poor households; 2) capacity (e.g. land, labor) to

engage in agricultural production; 3) a child attending the preschool with which they are about to sign the contract; and 4) signed agreement on a voluntary basis.

The aforementioned criteria imply that the treatment group smallholders included in the baseline survey have to withdraw from the programme when their children or grandchildren graduated from preschools. This led to high turnover rates of the beneficiaries. As a result, we are not able to construct a panel data for programme evaluation. This report is thus mainly based on a comparison between the treatment and comparison smallholders after the program interventions. Therefore, we conducted both qualitative focus group and case studies to complement the quantitative analysis based on the survey data. Meanwhile, to estimate the impact of the HGSF, a comparison of the changes in key indicators between treatment and comparison groups at the baseline and endline surveys was also conducted.

Focus Group Discussion

To better understand the impact of the HGSF on farmer households as well as underlying mechanisms, we organized a number of multi-stakeholder's participatory focus group discussions in selected preschools. Participants of these discussions include county program officers, preschool principals, and representative farmer households. Questions discussed include 1) contract details, including frequency of providing products to preschools, prices of the provided agricultural products, payment method, transportation of the provided agricultural products, merits and shortages of the contracting, 2) factors potentially impact contracting, especially concerns from preschools, 3) the impact of providing agricultural inputs and trainings on household food security, 4) impact on food loss and waste, 5) green production and food safety, and 6) impact on women empowerment.

Eight focus group discussions at Longshan county and twelve at Yongshun county were organized, and eight thematic analysis were drafted including value chain improvement, food provision, gaps between products provided by contracted household and demand from preschool, impact on diet diversity and production diversity, green production and food safety, impact on food loss and waste, barriers in delivering the products to preschools, and impact on women empowerment.

Key informant Interviews.

Given the fact that 83% (60 out of the 72) of the treatment group households who participated in the baseline survey were no longer supplying preschools at the time the endline survey was conducted, we organized in-depth key informant interviews with those treatment households who participated both the baseline and endline surveys (12 households) as well as local program officers. Information gathered from key informant interview is similar as in the focus group discussion but more targeted to each individual case.

Annex 8. The Impact of COVID-19

A8.1 COVID-19 Limitations on Project Implementation and Evaluation

1. COVID-19 and the related policy measures to curb the spread of the virus are expected to have considerable impacts on the school feeding intervention and accompanying evaluation. China was the first country to be hit by the COVID-19 pandemic which originated in the city of Wuhan, Hubei Province, where the first confirmed cases were reported in December 2019. Faced with the escalating virus outbreak, the Government implemented large-scale restrictions on mobility in late January 2020 in an attempt to limit COVID-19's spread. The main effect of COVID on project activities was that feeding activities were suspended from March to May 2020 as school closures were imposed nationwide. By the endline survey of the project, children in the treatment group had approximately two months less exposure to the project activities than was originally expected. This meant that during that time children lacked school meals that were directly contributing to their dietary intake and households' food security. Additionally, WFP suspended or canceled some nutrition education workshops, trainings, and meetings to avoid large gatherings during this period. However, local project office reported having moved some meetings and trainings online.

2. Early reports and evaluations indicate that children who relied on nutrition services provided by schools may suffer from worsening health and nutritional status in the short and medium term. Nutrition shocks, especially for the youngest children, are also expected to have strong long-term impacts on test scores, educational attainment, income, absenteeism and health (Almond and Currie, 2011; Sudfeld et al., 2015; Andrabi, Daniels and Das, 2020). Furthermore, lost schooling and learning may also lead to poor nutrition and health for themselves and their children in the long term (UNICEF, 2021). The UN reports that data show a deeply concerning trend: Since the beginning of the pandemic, UNESCO estimated that 1.6 billion learners in 199 countries worldwide were affected by school closures, with nearly 370 million children not receiving a school meal in 150 countries (UNESCO, 2020; WFP, 2020). In 2020, globally, an estimated 39 billion in-school meals have been missed during school closures. Children globally are estimated to have missed an average of 4 out of 10 in-school meals they would have regularly received, with children in some countries missing as many as 9 out of 10 in-school meals.

3. That said, the restrictions and intensity of the pandemic experience did have an unexpected impact on parenting experiences. Migrant parents stayed at home longer after the Spring Festival due to travel restrictions and thus spent more time with their children. Teachers in our study sites reported that children were livelier during this period.

4. It is also important to note the significant impact of COVID-19 on the evaluation process itself. Due to the budget constraint, the project did not carry out any evaluation activity at midline and limited documentation of participants' actions and food security and nutrition status during covid-19 should be acknowledged. Once Covid-19 related restrictions were lifted, and it was deemed sufficiently safe for research teams to travel to the field for research purposes, we managed to follow up the sample at the endline survey, which was postponed from June 2020 to May 2021. As data collection was largely conducted after the COVID-19 outbreak, respondents' attitudes, perceptions, and behaviors regarding project outcomes therefore have been affected by COVID-19. This means that it is possible that some potential effects of project activities could have been wearing off during this period.

5. Longer than planned time lag between the baseline and endline surveys also presents a unique risk as it became more difficulty to trace the children who participated in the baseline survey. Specifically, children aged 4-5 in the baseline would have graduated from preschools into primary schools, children aged younger than 4 years may have transferred to other preschools outside of our sample preschools, or sample counties even sample prefecture. Although we tried our best to trace those children by visiting them in their current schools/preschools to conduct face-to-face interviews, or called their primary caregivers to conduct interview by telephone, we have to face the reality that we might miss data on some modules for some students, or some sample children may even have attrited from the study. Either way, the data quality might be compromised during the survey postponement.

A8.2 Perceived Effects of COVID-19 Reported by Study Participants

6.   Although not an explicit area of focus for this evaluation, we added a limited number of Covid-19 related questions to most questionnaires in an effort to account for the effects of the pandemic and related restrictions on participants and their families. This section focuses on the findings that emerged from the data gathered on Covid-19, and we assessed the impact of COVID-19 on our sample to understand if it differentially impacted our treatment and comparison groups.

A8.2.1 Effects on employment and income

7.   Participants were asked about changes in both employment and income of household members as a consequence of the pandemic from 24 January to 11 March 2020, when large-scale social restrictions were still in place. As shown in Figure 8.1, a considerable amount of the individuals reported their job status and wages had deteriorated due to COVID-19, ranging from 14.9% of household members reporting a decline in monthly wages to 37.9% reporting being laid off. 14.5% of household members experienced job loss46. More than half of individuals who were back home for the Spring Festival experienced delays in returning to the cities where they used to work in late January. Generally, a larger share of working men than women in our sample were laid off or lost their jobs.
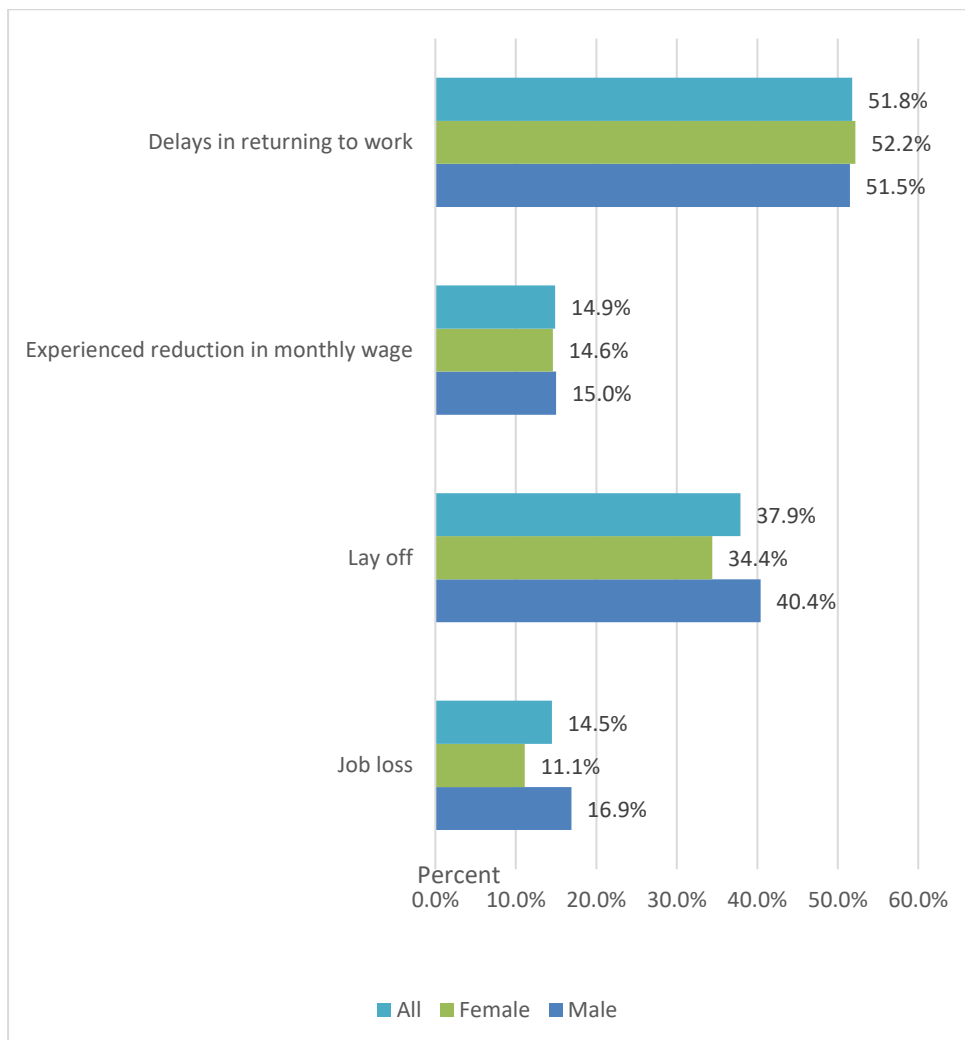
Figure A8.1 Percentage of individuals having experienced job or wage disruptions, by sex

8.   Consistent with analysis in previous sections, we construct household level indicators based on individual responses above to capture experiences of the households for each child in our study

---

46 Workers who said they experienced a job loss could have found a new job during that period, or they could have lost one of multiple jobs.

during the Covid-19. We then split the households into four mutually exclusive groups: those with both job and income loss (laid off or reduction in monthly wage), those with only job loss, those with only income loss, and those with neither job nor income loss. In order to represent this categorical variable, we created 3 dummy variables: "Job and income loss (9.6% of the sample)," "Only job loss (4.5% of the sample)," and "Only Income Loss (16.7% of the sample)." As Table 1 shows, we did not find any significant difference in the households of children in the treatment group versus those in the comparison group.

9.   It is important to underscore that job and income loss is at the household level. It is possible for a parent to report that he or she did not lose his or her job, but the household income nonetheless declined due to another householder's job loss. Or the parent might not have lost his or her job, but monthly wages may have declined, or he/she has been laid off, leading to an income loss. It's also possible that those who experienced job loss did not necessarily experience an income decline in the meantime, thanks to public policy interventions, etc. For these reasons it is important to distinguish children whose household experienced job and income losses jointly in the COVID-19 pandemic from those in which families experienced only one of these events.

Table A8.1 Household Experience with Job/Income Losses During Covid-19,

by treatment status

Variable	(1) Comparison group	(2) Treatment group	(3) Difference (2) – (1)
Job and Income Loss	0.09	0.10	0.01
	(0.29)	(0.30)	(0.02)
Only Job Loss	0.04	0.05	0.00
	(0.20)	(0.21)	(0.01)
Only Income Loss	0.17	0.16	-0.01
	(0.38)	(0.37)	(0.02)
Observations	599	475	1,074

Statistically significant at * $p<0.05$, **$p<0.01$, ***$p<0.001$

10.   We further examine whether the COVID-19 pandemic has had different effects on different household types. Figure 8.2 shows the percentage of households that reported job or income loss by their types. Overall, the effects are not very different between registered poor households and their counterparts. For example, 31% of poorer households have experienced a decline in income while the figure for the remaining households is about 30.6%. However, when it comes to ethnicity disparity, significantly more ethnic minority households (31.5%) reported to have experienced job or income loss than Han households (24.8%). Households with younger child (32.8%) seem to suffer more than those with older child (28.7%) during Covid-19. Also as expected, a larger share of households headed by a person with low or no literacy reported job/income loss than those households headed by a person with secondary or higher education.
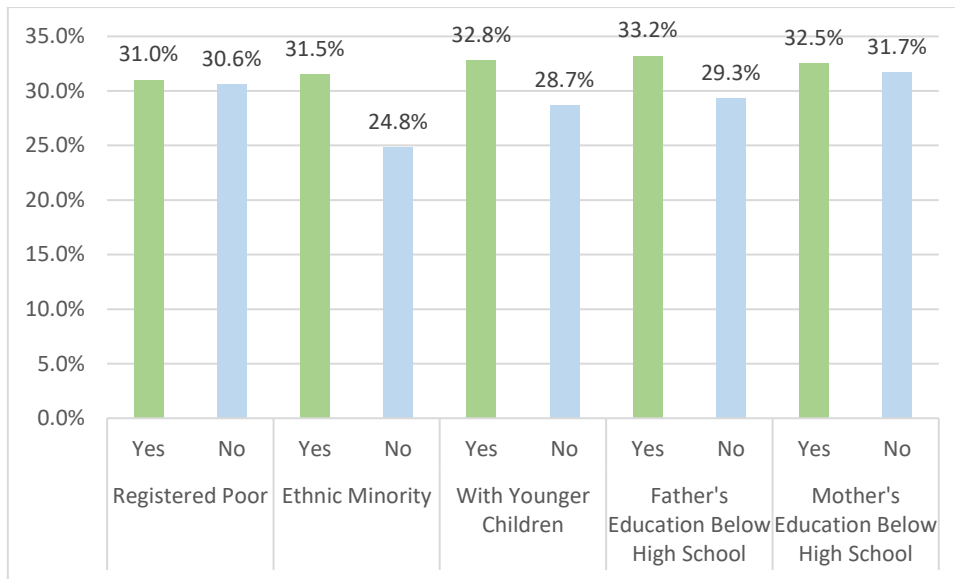
Figure A8.2 Households with job or income loss, by groups

A8.2.2 Effects on dietary change of households

11.   To understand the potential effects of COVID-19 on dietary diversity, our survey asked if the household's diet had changed due to COVID-19 compared to an average month before the COVID-19. However, Covid-19 and the lockdown measures made it difficult to collect the profiles during the pandemic, thus the information was collected in a retrospective method and thus may suffer from some recall bias.

12.   As shown in Figure 8.3, the vast majority of sample households reported that COVID-19 did not impact the food that they ate. The most frequent change was on reductions in meat and fish (24%), followed by reductions in snacks and organ meat (10%), due to price increase and difficulty buying or accessing food due to the state of emergency and shop closures.
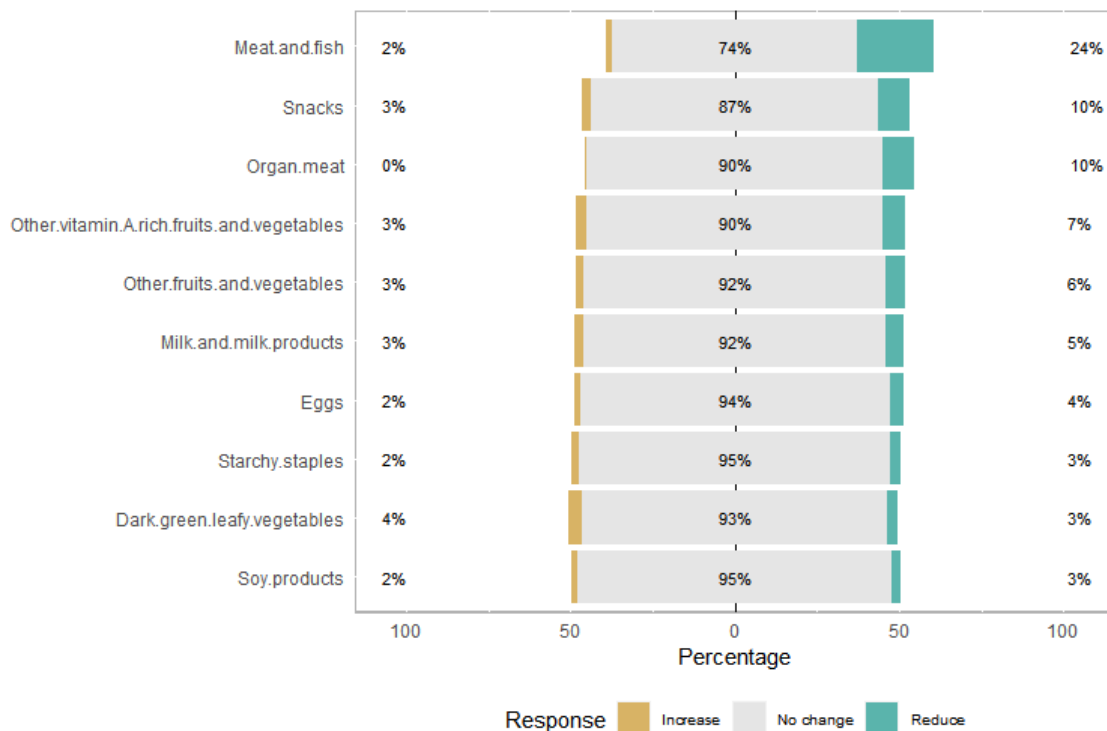
Figure A8.3 Dietary changes during Covid-19 across food groups

13.   Overall, we did not find differences in the impact of the Covid-19 pandemic on the diet of children from the treatment and comparison groups . This means that if we compare two groups with the same background and characteristics, who only differ in terms of treatment status, these children would be affected by the pandemic in a similar way.

A8.2.3 Other channels of effects

14.   Most schools were open before Covid-19 restrictions were imposed. There were only two exceptions in Yongshun county. Most schools re-opened once Covid-19 restrictions were lifted (and data collection could take place). School principals did not report any significant decrease in school attendance when their school re-opened, and the majority of study participants were expected to have returned to school. Thus, the impact of Covid-19 restrictions in terms of access to schooling and dropout rates is not significant in our study sites. However, in addition to economic pressures and dietary changes, school closures due to the Covid-19 pandemic did have limited the contact time between educators and children and disrupted curriculum coverage, and thus could exert a bias on children's cognitive abilities.

15.   Although not explicitly covered in our survey, there are also less obvious but equally severe social consequences, including greater risks of abuse and gender based violence. The psychosocial impact of COVID-19 is also becoming clearer, along with school closures and domestic lockdown cutting off access to peers, leisure activities and alternative sources of caregiving (WFP, FAO & UNICEF, 2020).

A8.3 Reflections

16.   It has proven challenging to ensure that school children who usually receive school feeding were picked up by social safety nets. Thus, the Covid-19 pandemic underscores the need for emergency preparedness and adaptation plans for school feeding. Consequently, many countries and international organizations have adapted their school feeding programs. According to World Food Programme's monitoring data on global school feeding programs, in responding to the absence of school-based provision of meals during school closures, a number of different modalities were employed to substitute the daily meal that children previously received at school. These alternatives include take-home rations, unconditional cash transfers and alternative hot meals. While a couple of countries reported putting programs on hold, which is similar to our case here (WFP, 2020), international efforts made under the Covid-19 pandemic to maintain the transfer even when schools were closed have generated some lessons: the project could look for alternative ways to support children and maintain the transfer even when schools are closed, while ensuring compliance with COVID-19 protocols. While evaluation of these alternative measures is under way, initial results suggest that they do provide benefits. With the reality of increasingly frequent shocks, it would be important to give more emphasis on preparedness and adaptation plans around shock-responsive school feeding programs in future situations where schools have to be closed.

Annex 9. Tables for Evaluation Results on Children's Nutrition and Development

Table A9.1. Baseline summary statistics (Balance checks)

	Balance check between treatment and comparison			Balance check between attrited and non-attrited		
	Treatment Group	Comparison Group	t-Test (p-value) H0: (1) = (2)	Attrited Sample	Non-Attrited Sample	t-Test (p-value) H0: (1) = (2)
	(1)	(2)	(3)	(4)	(5)	(6)
Outcome: Physical Indicators						
1)Hemoglobin Level (g/l)	115.96 [10.53]	115.73 [11.26]	0.81	116.49 [10.74]	115.68 [10.95]	0.24
2)Anemia	0.31 [0.46]	0.36 [0.48]	0.20	0.30 [0.46]	0.34 [0.47]	0.16
3)Height (cm)	102.86 [6.78]	102.58 [7.51]	0.82	102.22 [7.51]	102.83 [7.08]	0.35
4)HAZ	-0.81 [0.91]	-0.86 [1.01]	0.37	-0.80 [1.00]	-0.85 [0.96]	0.47
5)Stunted	0.10 [0.30]	0.11 [0.32]	0.44	0.09 [0.29]	0.11 [0.31]	0.38
6)Weight (kg)	16.28 [2.50]	16.20 [2.67]	0.83	16.13 [2.61]	16.27 [2.58]	0.56
7)WAZ	-0.53 [0.93]	-0.57 [0.88]	0.55	-0.50 [0.88]	-0.56 [0.91]	0.38
8)Underweight	0.06 [0.23]	0.05 [0.22]	0.54	0.05 [0.21]	0.06 [0.23]	0.50
9)Wasted	0.02 [0.14]	0.02 [0.12]	0.67	0.00 [0.06]	0.02 [0.14]	0.00***
10)Overweight	0.12 [0.33]	0.10 [0.30]	0.31	0.10 [0.31]	0.12 [0.32]	0.64
11)Obesity	0.03 [0.17]	0.02 [0.14]	0.31	0.03 [0.18]	0.02 [0.15]	0.34
Outcome: Cognitive Indicators						
12)VCI score	86.08 [12.39]	85.96 [13.08]	0.91	86.35 [13.26]	85.90 [12.57]	0.59
13)low VCI score	0.31 [0.46]	0.30 [0.46]	0.80	0.30 [0.46]	0.31 [0.46]	0.74
14)Extremely Low VCI score	0.09 [0.28]	0.12 [0.32]	0.25	0.10 [0.30]	0.10 [0.30]	0.86
15)WMI score	90.87 [12.60]	90.02 [14.20]	0.46	89.41 [13.85]	90.77 [13.31]	0.11
16)low WMI score	0.20 [0.40]	0.22 [0.42]	0.54	0.24 [0.43]	0.20 [0.40]	0.10
17)Extremely Low WMI score	0.03 [0.18]	0.06 [0.24]	0.09	0.06 [0.24]	0.05 [0.21]	0.40
Outcome: Socio-emotional Indicators						

	Balance check between treatment and comparison			Balance check between attrited and non-attrited		
	Treatment Group	Comparison Group	t-Test (p-value) H0: (1) = (2)	Attrited Sample	Non-Attrited Sample	t-Test (p-value) H0: (1) = (2)
	(1)	(2)	(3)	(4)	(5)	(6)
SDQ Subscale Scores						
18)Total Difficulties	12.56 [4.83]	12.21 [4.47]	0.28	12.79 [4.77]	12.21 [4.59]	0.04**
19)Emotional Symptoms	3.20 [2.11]	3.01 [1.92]	0.17	3.21 [2.03]	3.06 [2.01]	0.29
20)Conduct Problems	1.87 [1.62]	1.74 [1.45]	0.16	1.85 [1.54]	1.78 [1.53]	0.45
21)Hyperactivity	4.95 [2.23]	4.95 [2.15]	0.97	5.09 [2.34]	4.90 [2.12]	0.16
22)Peer Relationship Problems	2.53 [1.80]	2.52 [1.75]	0.94	2.64 [1.91]	2.48 [1.72]	0.14
23)Prosocial Behaviors	6.75 [2.13]	6.89 [2.25]	0.39	6.84 [2.26]	6.82 [2.17]	0.88
Abnormal Rates						
24)Total Difficulties	0.21 [0.41]	0.19 [0.39]	0.45	0.23 [0.42]	0.19 [0.39]	0.17
25)Emotional Symptoms	0.25 [0.43]	0.22 [0.42]	0.25	0.24 [0.43]	0.24 [0.42]	1.00
26)Conduct Problems	0.16 [0.37]	0.12 [0.33]	0.05	0.17 [0.37]	0.13 [0.34]	0.17
27)Hyperactivity	0.14 [0.35]	0.12 [0.32]	0.25	0.16 [0.37]	0.11 [0.32]	0.07*
28)Peer Relationship Problems	0.06 [0.25]	0.05 [0.22]	0.42	0.08 [0.27]	0.05 [0.22]	0.13
29)Prosocial Behaviors	0.15 [0.35]	0.16 [0.37]	0.60	0.16 [0.37]	0.15 [0.36]	0.47
30)ASQ:SE Abnormal Rates	0.70 [0.46]	0.69 [0.46]	0.72	0.70 [0.46]	0.69 [0.46]	0.97
Socio-Demographic Characteristics						
31)Boy	0.53 [0.50]	0.50 [0.50]	0.37	0.53 [0.50]	0.51 [0.50]	0.71
32)Elder children	0.49 [0.50]	0.49 [0.50]	0.97	0.42 [0.49]	0.51 [0.50]	0.08*
33)non-Han minority	0.93 [0.26]	0.85 [0.35]	0.02**	0.90 [0.31]	0.89 [0.32]	0.68
34)Left-behind children	0.73 [0.44]	0.69 [0.46]	0.13	0.73 [0.45]	0.70 [0.46]	0.47
35)Registered poor family	0.42 [0.49]	0.27 [0.44]	0.00***	0.31 [0.46]	0.35 [0.48]	0.28
36)Subsistence allowance family	0.11 [0.31]	0.07 [0.26]	0.14	0.13 [0.34]	0.08 [0.27]	0.05**
37)Longshan	0.42 [0.49]	0.47 [0.50]	0.65	0.27 [0.44]	0.49 [0.50]	0.00***
38)Father has at least senior high school degree	0.12 [0.32]	0.13 [0.34]	0.50	0.18 [0.39]	0.11 [0.32]	0.01***
39)Mother has at least senior high school degree	0.11 [0.31]	0.13 [0.34]	0.32	0.14 [0.35]	0.12 [0.32]	0.29

	Balance check between treatment and comparison			Balance check between attrited and non-attrited		
	Treatment Group	Comparison Group	t-Test (p-value) H0: (1) = (2)	Attrited Sample	Non-Attrited Sample	t-Test (p-value) H0: (1) = (2)
	(1)	(2)	(3)	(4)	(5)	(6)
Obs	637	697		260	1074	

Notes: a) Means with standard deviations reported in brackets. b) Cluster-robust standard errors adjusted for clustering at the class level in parentheses. c) *** p<0.01, ** p<0.05, * p<0.1

Table A9.2. Effects of the school feeding treatment on hemoglobin level and anemia

	OLS				IV			
	Hemoglobin (1)	Any anemia (2)	Mild anemia (3)	Moderate or severe anemia (4)	Hemoglobin (5)	Any anemia (6)	Mild anemia (7)	Moderate or severe anemia (8)
Treatment	-0.16	-0.01	-0.01	-0.01	-0.18	-0.02	-0.01	-0.01
	(0.63)	(0.01)	(0.01)	(0.01)	(0.71)	(0.02)	(0.01)	(0.01)
Hemoglobin at baseline	0.27***				0.27***			
	(0.03)				(0.03)			
Any anemia at baseline		0.09***				0.74***		
		(0.02)				(0.12)		
Mild anemia at baseline			0.06***				0.60***	
			(0.02)				(0.14)	
Moderate or severe anemia at baseline				0.05**				0.73***
				(0.02)				(0.21)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,045	1,046	1,046	1,046	1,045	1,046	1,046	1,046
R-squared	0.11				0.11	0.05	0.02	0.02
Pseudo R-squared		0.0953	0.0658	0.0687				

Notes: Coefficients are estimated separately in each column. Cluster-robust standard errors adjusted for clustering at the class level in parentheses. Any anemia is defined as a hemoglobin value < 11.5 g/dl, mild anemia is defined as a hemoglobin value ≥ 11 & < 11.5 g/dl, moderate/severe anemia is defined as a hemoglobin value < 11 g/dl. All regressions control for baseline value of the dependent variable, additional controls include child's gender, age, ethnicity, household registered poverty, father's education, and the migration status of the child's parents at baseline.*** p<0.01, ** p<0.05, * p<0.1

Table A9.3. Heterogeneous effects of the school feeding treatment on hemoglobin, by baseline hemoglobin level

	OLS				IV			
	Baseline Hb > 125 g/L (1)	Baseline 115 g/L < Hb < 125 g/L (2)	Baseline 110 g/L < Hb < 115 g/L (3)	Baseline Hb < 110 g/L (4)	Baseline Hb > 125 g/L (5)	Baseline 115 g/L < Hb < 125 g/L (6)	Baseline 110 g/L < Hb < 115 g/L (7)	Baseline Hb < 110 g/L (8)
Treatment	0.09	-0.92	0.46	1.20	0.09	-1.03	0.55	1.45
	(1.72)	(0.93)	(1.39)	(1.17)	(1.74)	(1.02)	(1.61)	(1.43)
Hemoglobin at baseline	0.34**	0.45***	-0.06	0.10	0.34**	0.45***	-0.05	0.10
	(0.16)	(0.15)	(0.39)	(0.08)	(0.16)	(0.15)	(0.38)	(0.08)
Constant	88.62***	78.35***	136.63***	118.77***	88.60***	78.48***	136.42***	118.48***
	(21.28)	(18.49)	(43.84)	(8.85)	(20.64)	(18.27)	(42.39)	(8.73)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	197	404	157	284	197	404	157	284
R-squared	0.08	0.04	0.06	0.02	0.08	0.04	0.06	0.02

Notes: Coefficients are estimated separately in each column. Cluster-robust standard errors adjusted for clustering at the class level in parentheses. All regressions control for baseline value of the dependent variable, additional controls include child's gender, age, ethnicity, household registered poverty, father's education, and the migration status of the child's parents at baseline.*** p<0.01, ** p<0.05, * p<0.1

Table A9.4. Heterogeneous effects of the school feeding treatment on anemia, by gender and age

	OLS			IV		
	Any (1)	Mild (2)	Moderate (3)	Any (4)	Mild (5)	Moderate (6)
Panel A. Girls						
Treatment	-0.02	-0.02	-0.00	-0.03	-0.02	-0.00
	(0.02)	(0.02)	(0.01)	(0.03)	(0.02)	(0.02)
Any anemia at baseline	0.10***			0.11***		
	(0.03)			(0.03)		
Mild anemia at baseline		0.04*			0.05*	
		(0.02)			(0.03)	
Moderate or severe anemia at baseline			0.11*			0.10*
			(0.06)			(0.05)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	506	506	460	506	506	506
R-squared				0.06	0.02	0.04
Pseudo R-squared	0.122	0.0732	0.128			
Girls-Age 3 at baseline						
Treatment	-0.08*	-0.06*	-0.04	-0.11*	-0.09**	-0.03
	(0.05)	(0.03)	(0.04)	(0.06)	(0.04)	(0.03)
Any anemia at baseline	0.10*			0.12**		
	(0.06)			(0.06)		
Mild anemia at baseline		0.09*			0.11*	
		(0.05)			(0.06)	
Moderate or severe anemia at baseline			0.06			0.05
			(0.11)			(0.07)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	185	185	121	185	185	185
R-squared				0.07	0.05	0.03
Pseudo R-squared	0.161	0.160	0.0725			
Girls-Age 5 at baseline						
Treatment	-0.00	0.00	0.00	-0.01	0.00	-0.00
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)
Any anemia at baseline	0.05*			0.05*		
	(0.03)			(0.03)		

	(1)	(2)	(3)	(4)	(5)	(6)
Mild anemia at baseline		0.00			0.00	
		(0.03)			(0.03)	
Moderate or severe anemia at baseline			0.08			0.10
			(0.06)			(0.07)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	240	240	214	267	267	267
R-squared				0.03	0.02	0.07
Pseudo R-squared	0.0634	0.0706	0.250			
Panel B. Boys						
Treatment	-0.01	0.00	-0.02	-0.01	0.01	-0.02
	(0.02)	(0.01)	(0.02)	(0.03)	(0.02)	(0.02)
Any anemia at baseline	0.11***			0.12***		
	(0.03)			(0.03)		
Mild anemia at baseline		0.08***			0.08***	
		(0.03)			(0.03)	
Moderate or severe anemia at baseline			0.05			0.05
			(0.04)			(0.04)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	540	540	484	540	540	540
R-squared				0.05	0.04	0.01
Pseudo R-squared	0.0990	0.0909	0.0355			
Boys-Age 3 at baseline						
Treatment	0.04	0.04	-0.00	0.05	0.05	0.00
	(0.04)	(0.03)	(0.02)	(0.04)	(0.03)	(0.02)
Any anemia at baseline	0.14***			0.13***		
	(0.05)			(0.05)		
Mild anemia at baseline		0.08*			0.08*	
		(0.05)			(0.04)	
Moderate or severe anemia at baseline			0.11			0.08
			(0.08)			(0.06)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	201	201	173	231	231	231
R-squared				0.09	0.05	0.03
Pseudo R-squared	0.112	0.0657	0.0868			
Boys-Age 5 at baseline						
Treatment	-0.04*	-0.02	-0.03	-0.05	-0.03	-0.03

	(0.02)	(0.01)	(0.03)	(0.03)	(0.02)	(0.03)
Any anemia at baseline	0.11**			0.11**		
	(0.04)			(0.05)		
Mild anemia at baseline		0.08*			0.08*	
		(0.04)			(0.04)	
Moderate or severe anemia at baseline						-0.02*
						(0.01)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	251	251	191	251	251	251
R-squared				0.06	0.05	0.01
Pseudo R-squared	0.136	0.188	0.0331			

Notes: Coefficients are estimated separately in each column. Cluster-robust standard errors adjusted for clustering at the class level in parentheses. All regressions control for baseline value of the dependent variable, additional controls include child's gender, age, ethnicity, household registered poverty, father's education, and the migration status of the child's parents at baseline.*** p<0.01, ** p<0.05, * p<0.1

Table A9.5. Effects of the school feeding treatment on anthropometry indexes

OLS

Panel A	HAZ (1)	Stunting (2)	WAZ (3)	Underweight (4)	Wasting (5)	Overweight (6)
Treatment	0.01	-0.01	0.03	0.01	-0.00	0.04*
	(0.06)	(0.01)	(0.04)	(0.01)	(0.01)	(0.02)
Baseline level	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,048	1,048	1,050	1,050	1,048	1,048
R-squared	0.62		0.69			

IV

Panel B	HAZ (1)	Stunting (2)	WAZ (3)	Underweight (4)	Wasting (5)	Overweight (6)
Treatment	0.01	-0.02	0.03	0.01	-0.00	0.04*
	(0.06)	(0.01)	(0.05)	(0.01)	(0.01)	(0.02)
Baseline level	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,048	1,048	1,050	1,050	1,048	1,048
R-squared	0.62	0.34	0.69	0.41	0.05	0.20
Pseudo R-squared		0.454		0.462	0.0781	0.173

Notes: Cluster-robust standard errors adjusted for clustering at the class level in parentheses. All regressions control for baseline value of the dependent variable, additional controls include child's gender, age, ethnicity, household registered poverty, father's education, and the migration status of the child's parents at baseline.*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A9.6. Heterogeneous effects of the school feeding treatment on anthropometry indexes (ITT), by gender and age

Panel A. Girls	HAZ (1)	WAZ (2)	Stunting (3)	Underweight (4)	Wasting (5)	Overweight (6)
Treatment	-0.00	-0.06	0.00	0.01	-0.00	0.03
	(0.06)	(0.05)	(0.01)	(0.01)	(0.01)	(0.03)
Baseline level	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	508	509	508	509	508	508
R-squared	0.59	0.73				
Pseudo R-squared			0.584	0.587	0.109	0.124
Girls-Age 3 at baseline						
Treatment	0.04	0.03	0.00	-0.01	-0.04	0.04*
	(0.11)	(0.08)	(0.01)	(0.03)	(0.02)	(0.02)
Baseline level	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	185	186	159	80	154	185
R-squared	0.44	0.70				
Pseudo R-squared			0.397	0.218	0.0452	0.265
Girls-Age 5 at baseline						
Treatment	-0.05	-0.10	0.00	0.03*	0.01	-0.00
	(0.08)	(0.07)	(0.01)	(0.02)	(0.02)	(0.05)
Baseline level	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	269	269	269	269	241	269
R-squared	0.72	0.77				
Pseudo R-squared			0.651	0.668	0.325	0.112
Panel B. Boys	HAZ (1)	WAZ (2)	Stunting (3)	Underweight (4)	Wasting (5)	Overweight (6)
Treatment	0.04	0.15**	-0.02*	0.00	-0.00	0.06*
	(0.07)	(0.06)	(0.01)	(0.01)	(0.01)	(0.04)
Baseline level	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	540	541	540	541	540	540
R-squared	0.65	0.67				
Pseudo R-squared			0.376	0.386	0.0788	0.187

Boys-Age 3 at baseline						
Treatment	0.17	0.29***	-0.01	-0.01	-0.03	0.12*
	(0.10)	(0.10)	(0.01)	(0.02)	(0.02)	(0.07)
Baseline level	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	231	231	231	201	231	231
R-squared	0.57	0.65				
Pseudo R-squared			0.451	0.297	0.125	0.221
Boys-Age 5 at baseline						
Treatment	-0.07	-0.05	-0.03	0.03**	0.03	-0.00
	(0.11)	(0.09)	(0.03)	(0.01)	(0.02)	(0.05)
Baseline level	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	251	252	233	218	217	251
R-squared	0.72	0.70				
Pseudo R-squared			0.310	0.550	0.180	0.153

Table A9.7. Heterogeneous effects of the school feeding treatment on anthropometry indexes (ITT), by household poverty

	HAZ	WAZ	Stunting	Underweight	Wasting	Overweight
Panel A. Poverty households	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.06	0.06	-0.00	0.01	-0.04**	0.06**
	(0.09)	(0.05)	(0.01)	(0.01)	(0.02)	(0.03)
Baseline level	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	392	393	392	369	392	392
R-squared	0.62	0.74				
Pseudo R-squared			0.520	0.458	0.153	0.204
Panel B. Non-poverty households						
Treatment	0.07	0.02	-0.01	0.00	0.02	0.03
	(0.06)	(0.05)	(0.01)	(0.01)	(0.01)	(0.03)
Baseline level	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	656	657	656	657	656	656
R-squared	0.62	0.66				
Pseudo R-squared			0.423	0.487	0.0619	0.164

Notes: Coefficients are estimated separately in each column. Cluster-robust standard errors adjusted for clustering at the class level in parentheses. Household poverty is defined either being registered poor or receiving subsistence allowance at baseline. All regressions control for baseline value of the dependent variable, additional controls include child's gender, age, ethnicity, father's education, and the migration status of the child's parents at baseline.*** p<0.01, ** p<0.05, * p<0.

Table A9.8. Effects of the school feeding treatment on VCI and WMI scores

	OLS/Probit						IV					
	ΔVCI	Low VCI Scores	Extremely Low VCI Scores	ΔWMI	Low WMI Scores	Extremely Low WMI Scores	ΔVCI	Low VCI Scores	Extremely Low VCI Scores	ΔWMI	Low WMI Scores	Extremely Low WMI Scores
	(1)	(2)	(3)	(4)	(5)	(6)						
Treatment Group	-1.12	-0.04	0.02	-0.35	0.01	0.00	-1.07	-0.03	0.01*	0.15	0.01	0.00
	(1.01)	(0.03)	(0.01)	(1.01)	(0.03)	(0.01)	(1.05)	(0.03)	(0.01)	(1.03)	(0.03)	(0.01)
Control Variables	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Baseline Level	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Constant	57.44***			79.68***			57.38***			79.64***		
	(3.56)			(2.99)			(3.56)			(2.98)		
Observations	966	966	966	972	972	863	966	966	966	972	972	863
R-squared	0.35			0.56			0.35			0.56		
Pseudo R-squared		0.04	0.15		0.01	0.07		0.04	0.18		0.01	0.09

Notes: a) Cluster-robust standard errors adjusted for clustering at the class level in parentheses. b) Marginal effects are shown for Probit models. c) *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A9.9. Heterogeneous effects of the school feeding treatment on cognitive level (ITT), by gender, age and poverty status

	ΔVCI	Low VCI Scores	ΔWMI	Low WMI Scores
Panel A. Girls	(1)	(2)	(3)	(4)
Treatment	-1.21	-0.02	-1.24	0.01
	(1.11)	(0.04)	(1.25)	(0.04)
Baseline level	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	469	469	473	473
R-squared	0.34		0.55	
Pseudo R-squared		0.06		0.02
	ΔVCI	Low VCI Scores	ΔWMI	Low WMI Scores
Panel B. Boys	(1)	(2)	(3)	(4)
Treatment	-0.98	-0.06	0.54	0.01
	(1.51)	(0.04)	(1.38)	(0.04)
Baseline level	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	497	497	499	499
R-squared	0.36		0.58	
Pseudo R-squared		0.02		0.03
	ΔVCI	Low VCI Scores	ΔWMI	Low WMI Scores
Panel C. Younger Children	(1)	(2)	(3)	(4)
Treatment	-0.05	-0.04	-0.16	0.02
	(1.03)	(0.04)	(1.60)	(0.04)
Baseline level	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	448	448	445	445
R-squared	0.56		0.45	
Pseudo R-squared		0.04		0.01

	ΔVCI	Low VCI Scores	ΔWMI	Low WMI Scores
Panel D. Elder Children	(1)	(2)	(3)	(4)
Treatment	-1.91	-0.04	-0.50	0.01
	(1.55)	(0.04)	(1.24)	(0.04)
Baseline level	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	518	518	527	527
R-squared	0.13		0.44	
Pseudo R-squared		0.02		0.01
Panel E. Poverty households	ΔVCI	Low VCI Scores	ΔWMI	Low WMI Scores
	(1)	(2)	(3)	(4)
Treatment	**-3.70*****	-0.06	-1.46	0.03
	(1.25)	(0.04)	(1.47)	(0.04)
Baseline level	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	371	371	372	372
R-squared	0.34		0.55	
Pseudo R-squared		0.04		0.03
Panel F. Non-poverty households	ΔVCI	Low VCI Scores	ΔWMI	Low WMI Scores
	(1)	(2)	(3)	(4)
Treatment	0.41	-0.02	0.15	0.00
	(1.21)	(0.03)	(1.16)	(0.04)
Baseline level	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	595	595	600	600
R-squared	0.37		0.57	
Pseudo R-squared		0.04		0.01

Table A9.10. Heterogeneous effects of the school feeding treatment on cognitive level (ITT), by wealth index

Panel A. Quintile 1	ΔVCI (1)	Low VCI Scores (2)	ΔWMI (3)	Low WMI Scores (4)
Treatment	-3.25*	-0.06	-2.86	0.13**
	(1.73)	(0.06)	(1.84)	(0.06)
Baseline level	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	211	211	208	208
R-squared	0.31		0.53	
Pseudo R-squared		0.03		0.04
Panel B. Quintile 2	ΔVCI (1)	Low VCI Scores (2)	ΔWMI (3)	Low WMI Scores (4)
Treatment	-0.38	-0.04	1.17	0.04
	(1.98)	(0.06)	(1.79)	(0.06)
Baseline level	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	182	182	182	182
R-squared	0.37		0.59	
Pseudo R-squared		0.04		0.03
Panel C. Quintile 3	ΔVCI (1)	Low VCI Scores (2)	ΔWMI (3)	Low WMI Scores (4)
Treatment	-2.14	-0.03	-0.05	-0.00
	(2.01)	(0.07)	(1.67)	(0.06)
Baseline level	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	190	190	194	194
R-squared	0.30		0.57	
Pseudo R-squared		0.07		0.03

	ΔVCI	Low VCI Scores	ΔWMI	Low WMI Scores
Panel D. Quintile 4	(1)	(2)	(3)	(4)
Treatment	0.72	0.04	-0.42	-0.10*
	(1.66)	(0.07)	(2.13)	(0.06)
Baseline level	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	200	200	203	203
R-squared	0.44		0.61	
Pseudo R-squared		0.04		0.05
	ΔVCI	Low VCI Scores	ΔWMI	Low WMI Scores
Panel E. Quintile 5	(1)	(2)	(3)	(4)
Treatment	0.02	-0.12**	2.35	-0.04
	(2.27)	(0.06)	(1.89)	(0.06)
Baseline level	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	183	183	185	185
R-squared	0.36		0.57	
Pseudo R-squared		0.07		0.04

Table A9.11. Effects of the school feeding treatment on SDQ subscale scores

	OLS						IV					
	ΔTotal Difficulties	ΔEmotional Symptoms	ΔConduct Problems	ΔHyperactivity	ΔPeer Relationship Problems	ΔProsocial Behaviors	ΔTotal Difficulties	ΔEmotional Symptoms	ΔConduct Problems	ΔHyperactivity	ΔPeer Relationship Problems	ΔProsocial Behaviors
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treatment Group	-0.28	-0.17	-0.00	-0.07	-0.02	-0.07	-0.32	-0.19	-0.00	-0.08	-0.02	-0.08
	(0.35)	(0.13)	(0.09)	(0.17)	(0.10)	(0.13)	(0.39)	(0.14)	(0.10)	(0.19)	(0.12)	(0.15)
Control Variables	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Baseline Level	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Constant	7.46***	2.21***	1.11***	3.38***	1.44***	7.06***	7.50***	2.24***	1.11***	3.39***	1.44***	7.07***
	(0.77)	(0.23)	(0.18)	(0.32)	(0.23)	(0.38)	(0.77)	(0.23)	(0.18)	(0.33)	(0.22)	(0.37)
Observations	957	957	957	957	957	957	957	957	957	957	957	957
R-squared	0.36	0.45	0.48	0.39	0.45	0.49	0.36	0.45	0.48	0.39	0.45	0.49

Notes: a) Cluster-robust standard errors adjusted for clustering at the class level in parentheses. b) *** p<0.01, ** p<0.05, * p<0.1

Table A9.12. Effects of the school feeding treatment on SDQ abnormal outcomes

	Probit						IV					
	Total Difficulties	Emotional Symptoms	Conduct Problems	Hyperactivity	Peer Relationship Problems	Prosocial Behaviors	Total Difficulties	Emotional Symptoms	Conduct Problems	Hyperactivity	Peer Relationship Problems	Prosocial Behaviors
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treatment Group	-0.02	-0.03	-0.01	-0.04**	0.01	0.01	-0.02	-0.03	-0.01	-0.04**	0.01	0.01
	(0.02)	(0.03)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.01)	(0.02)
Control Variables	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Baseline Level	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	957	957	957	957	957	957	957	957	957	957	957	957
Constant							0.08*	0.11**	0.06*	0.09**	0.00	0.07*
							(0.05)	(0.07)	(0.03)	(0.04)	(0.02)	(0.04)
Pseudo R-squared	0.03	0.02	0.01	0.04	0.03	0.01						
R-squared							0.02	0.02	0.01	0.02	0.01	0.00

Notes: a) Cluster-robust standard errors adjusted for clustering at the class level in parentheses. b) Marginal effects are shown for Probit models. c) *** p<0.01, ** p<0.05, * p<0.1

Table A9.13. Effects of the school feeding treatment on ASQ:SE abnormal outcomes

	Abnormal ASQ:SE scores	
	Probit	IV
	(1)	(2)
Treatment Group	-0.07	-0.03
	(0.12)	(0.05)
Control Variables	YES	Yes
Baseline Level	YES	Yes
Constant		0.32***
		(0.06)
Observations	953	953
R-squared		
Pseudo R-squared	0.00	0.00

Notes: a) Cluster-robust standard errors adjusted for clustering at the class level in parentheses. b) Marginal effects are shown for Probit models. c) *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A9.14. Heterogeneous effects of the school feeding treatment on abnormal social-developmental status (ITT), by gender, age and poverty status

	Total Difficulties	Emotional Symptoms	Conduct Problems	Hyperactivity	Peer Relationship	Prosocial Behaviors	ASQ:SE
Panel A. Girls	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	-0.03	-0.01	0.00	-0.03	0.01	0.03	-0.04
	(0.02)	(0.04)	(0.02)	(0.02)	(0.02)	(0.03)	(0.04)
Baseline level	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES
Observations	470	470	470	470	393	470	469
Pseudo R-squared	0.08	0.02	0.03	0.08	0.03	0.01	0.01
	Total Difficulties	Emotional Symptoms	Conduct Problems	Hyperactivity	Peer Relationship	Prosocial Behaviors	ASQ:SE
Panel B. Boys	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.01	-0.05	-0.01	-0.04	0.02	-0.01	-0.01
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.05)
Baseline level	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES
Observations	487	487	487	487	487	487	484
Pseudo R-squared	0.01	0.03	0.01	0.03	0.07	0.02	0.01
	Total Difficulties	Emotional Symptoms	Conduct Problems	Hyperactivity	Peer Relationship	Prosocial Behaviors	ASQ:SE
Panel C. Younger Children	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	-0.07**	-0.05	-0.03	-0.06**	-0.01	0.01	-0.04
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.05)
Baseline level	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES
Observations	501	501	501	501	434	501	499
Pseudo R-squared	0.03	0.03	0.01	0.04	0.03	0.02	0.01

	Total Difficulties	Emotional Symptoms	Conduct Problems	Hyperactivity	Peer Relationship	Prosocial Behaviors	ASQ:SE
Panel D. Elder Children	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.03	-0.01	0.02	-0.01	0.03**	0.00	-0.01
	(0.02)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.05)
Baseline level	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES
Observations	456	456	456	456	456	456	454
Pseudo R-squared	0.09	0.03	0.04	0.09	0.10	0.01	0.02
	Total Difficulties	Emotional Symptoms	Conduct Problems	Hyperactivity	Peer Relationship	Prosocial Behaviors	ASQ:SE
Panel E. Poverty households	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	-0.03	-0.07*	0.01	-0.05*	0.03*	0.02	-0.02
	(0.03)	(0.04)	(0.03)	(0.03)	(0.02)	(0.03)	(0.06)
Baseline level	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES
Observations	358	358	358	358	251	358	356
Pseudo R-squared	0.06	0.05	0.03	0.08	0.09	0.05	0.04
	Total Difficulties	Emotional Symptoms	Conduct Problems	Hyperactivity	Peer Relationship	Prosocial Behaviors	ASQ:SE
Panel F. Non-poverty households	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	-0.01	-0.01	-0.01	-0.03	0.00	-0.00	-0.03
	(0.02)	(0.03)	(0.02)	(0.02)	(0.01)	(0.02)	(0.04)
Baseline level	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES
Observations	599	599	599	599	599	599	597
Pseudo R-squared	0.04	0.03	0.04	0.04	0.02	0.02	0.01

Table A9.15. Nine Food Groups of Dietary Diversity Scores (DDS)

Food group	Examples
Starchy staples	Cereals (corn/maize, rice, wheat, sorghum, millet or any other grains or foods made from these (e.g. bread, noodles, porridge or other grain products) + insert local foods e.g. ugali, nshima, porridge or paste) and white tubers and roots (white potatoes, white yam, white cassava, or other foods made from roots)
Dark green leafy vegetables	dark green/leafy vegetables, including wild ones + locally available vitamin A rich leaves such as amaranth, cassava leaves, kale, spinach etc.
Other vitamin A rich fruits and vegetables	pumpkin, carrot, squash, or sweet potato that are orange inside + other locally available vitamin A rich vegetables (e.g. red sweet pepper), ripe mango, cantaloupe, apricot (fresh or dried), ripe papaya, dried peach, and 100% fruit juice made from these + other locally available vitamin A rich fruits
Other fruits and vegetables	other vegetables (e.g. tomato, onion, eggplant) + other locally available vegetables, other fruits, including wild fruits and 100% fruit juice made from these
Organ meat	liver, kidney, heart or other organ meats or blood-based foods
Meat and fish	beef, pork, lamb, goat, rabbit, game, chicken, duck, other birds, insects, fresh or dried fish or shellfish
Eggs	eggs from chicken, duck, guinea fowl or any other egg
Legumes, nuts and seeds	dried beans, dried peas, lentils, nuts, seeds or foods made from these (e.g. hummus, peanut butter)
Milk and milk products	milk, cheese, yogurt or other milk products

Table A9.16. Heterogeneous effects of the school feeding treatment (ITT) on Nutrition Knowledge Score, by wealth index

	ΔNutrition Knowledge Score				
	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Treatment	-1.48	2.00	2.93	-0.70	5.31*
	(2.08)	(2.18)	(1.89)	(2.32)	(2.73)
Baseline level	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES
Observations	201	180	188	191	199
R-squared	0.37	0.32	0.39	0.35	0.29

Notes: a) Cluster-robust standard errors adjusted for clustering at the class level in parentheses. b) *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Annex 10. Characteristics of the Contracted Farmers and the Contract Enforcement Practices

Annex 10.1. Socio-demographic characteristics of sampled smallholders

1.   The data show that smallholders in the treatment and comparison groups are qualitatively similar in most observable socio-demographic characteristics. The only exception is those in the treatment group are less likely to be engaged in off-farm jobs than those in the comparison group. The comparison of the changes between treatment and comparison group from the baseline to the endline surveys shows the same feature.

Table A10.1 Socio-demographic characteristics of sampled smallholders, by treatment status

		Total sample (1)	Treatment group (2)	Comparison group (3)	Difference in means (4)=(2)-(3)
Personal characteristics					
Observations		665[a]	378	287	
Gender	male=1	0.538	0.542	0.533	0.009
Age	years	32.623	32.444	32.860	-0.416
Laborforce	aged between 14-65 =1	0.561	0.574	0.544	0.031
Education	years	4.577	4.447	4.749	-0.302
On-farm employment	on-farm employment =1	0.173	0.196	0.143	0.053*
Off-farm employment	off-farm employment =1	0.314	0.315	0.314	0.001
Self rated health status	healthy=1	0.759	0.749	0.774	-0.025
Household head characteristics					
Observations		106	58	48	
Gender	male=1	0.905	0.914	0.894	0.020
Age	years	53.505	53.069	54.043	-0.974
Education	years	5.752	5.672	5.851	-0.179
On-farm employment	on-farm employment =1	0.509	0.603	0.396	0.208**
Off-farm employment	off-farm employment =1	0.349	0.293	0.417	-0.124
Self rated health status	healthy=1	0.519	0.517	0.521	-0.004
Household characteristics					
Observations		106	58	48	
Household size	number	6.333	6.517	6.106	0.411
Gender	proportion of male members in each household	0.544	0.547	0.540	0.006
Laborforce	proportion of laborforce aged	0.556	0.570	0.539	0.031

		Total sample (1)	Treatment group (2)	Comparison group (3)	Difference in means (4)=(2)-(3)
Education_primary	between 14-65 =1 in each household proportion of household members completing primary school in each household	0.404	0.391	0.419	-0.027
Education_middle school	proportion of household members completing middle school in each household	0.244	0.241	0.248	-0.007
Education_high school	proportion of household members completing high school in each household	0.063	0.069	0.055	0.014
Education_college	proportion of household members completing college in each household	0.012	0.009	0.016	-0.007
On-farm employment	proportion of household members engaging in on-farm jobs in each household	0.184	0.208	0.154	0.054
Off-farm employment (outside)	proportion of household members engaging in off-farm (outside) jobs in each household	0.244	0.259	0.227	0.032
Off-farm employment (local)	proportion of household members engaging in off-farm (local) jobs in each household	0.059	0.042	0.080	-0.037**
Student proportion	proportion of household members who are currently at school in each household	0.161	0.150	0.173	-0.023
Preschooler proportion	proportion of preschoolers in each household	0.155	0.169	0.137	0.031

Notes: the observations is based on the count of all the household members of the 106 households

Annexes 10.2. Contract enforcement

2.   In practice, the demand for agricultural products was specified via the three-party agreement. The preschools agree to buy those foodstuffs produced by contract farmers at market price provided that certain food quality and safety standards are met. The program office is responsible for providing technical trainings and some production inputs such as seeds and fertilizer. The preschools' demand for foodstuffs is specified in the agreement: 1) the items of agricultural products to be supplied by the smallholders; 2) the delivery method. In the case of Xiangxi prefecture, smallholders are required to deliver foodstuffs to the preschool

directly; 3) the quality of the produces; 4) the quantity, aiming to purchase at least 30% of the foodstuffs for school meals from local small-scale farmers; and 5) payment and pricing mechanism.

• *Contract products*

3.   Agricultural products covered in the contracts include vegetables, poultry, eggs, cereal and oil, fruits and livestock products. The data show that 94% of the contracted households provided vegetables to their contracting preschools, followed by poultry at 31%.

Table A10.2 Types of agricultural produce provided by farmers under the contract

Contract items	Number of contract farmers	Percentage of contract farmers (%)
Vegetables	55	94
Poultry (chicken, duck)	18	31
Eggs	15	25
Cereals, oils	10	17
Fruits	3	6
Pork	1	2

Source: Field survey.

• *Price*

4.   The majority of the foodstuffs supplied to preschool is sold at market price. While 6% of contract farmers reported a higher than market price 6% reported a lower than market price.

Table A10.3 Price relative to market price of foodstuffs supplied to preschools

	Number of contract farmers	Percentage of contract farmers (%)
Higher than market price	3	6
At market price	51	88
Lower than market price	3	6

Source: Field survey.

• *Frequency*

5.   The data show that 44% of contract farmers supplied food on a weekly basis while 46% provided foodstuff on a monthly basis. 8% supplied at a frequency of once every 1-3 months. One contract farmer supplied at a much lower frequency, only at a quarterly basis.

Table A10.4 Transaction frequency

	Number of contract farmers	Percentage of contract farmers (%)
Within a week	26	44
Within a month	27	46
Within a quarter	5	8
Over a quarter and within half a year	1	2

Source: Field survey.

• *Difficulty in transportation*

6.   The data also show that 4% reported that it was extremely difficult for them to deliver food to the contracting preschools and 13% reported that it was difficult. By contrast, 59% of the contracted households reported they had no difficulty in delivering agricultural produce to preschool.

Table A10.5 Transportation difficulty

	Number of contract farmers	Percentage of contract farmers (%)
Very difficult	2	4

Difficult	8	13
Moderate	13	23
Easy	27	46
Very easy	8	13

Source: Field survey.

- *Paying method*

7.   98% of transactions were done via cash payment. The majority of the farmer households were paid on site when the transaction was made (94%) while 6% reported that it was paid at an agreed date.

Table A10.6 Payment method

	Number of contract farmers	Percentage of contract farmers (%)
Cash	57	98
Wechat	1	2
Paid on site	55	94
Paid at an agreed date	3	6

Source: Field survey.

- *Quantity and quality*

8.   In the interview, smallholders were asked to evaluate the enforcement of the contract. 92% of contract farmers can provide all kinds of produces as indicated in the contract. In the meantime, 85% of them met the quantity standard while the share of those satisfying the quality standard is 100%.

Table A10.7 Transaction quantity and quality

	Number of contract farmers	Percentage of contract farmers (%)
Satisfy the species requirement	53	92
Satisfy the quantity requirement	49	85
Satisfy the quality requirement	58	100

Source: Field survey.

- *Self-rated satisfaction towards the programme*

9.   On the whole, a majority of smallholders are satisfied with this program. In the survey, we also asked smallholders about their satisfaction with the smallholder assistance program. The share of satisfied and very satisfied householders accounted for 58% and 38%, respectively. The rest 4 percent reported that they are moderately satisfied.

Table A10.8 Self-rated satisfaction towards the programme

	Number of contract farmers	Percentage of contract farmers (%)
Very satisfied	22	38
Satisfied	34	58
Moderate	2	4

Source: Field survey.

Annex 11. Summary Terms of Reference

The terms of reference (TOR) for the "Preschool Nutrition Pilot in Selected Counties of China from Feburary 2018 to January 2021: An Evaluation from Baseline to Impact" are available on WFP's website using the following link: https://docs.wfp.org/api/documents/WFP-0000063838/download/?_ga=2.187477864.1384108412.1636622193-812731340.1564991010.

Annex 12. Bibliography

Abiodun T. Atoloye, Mateja R. Savoie-Roskos, Patricia M. Guenther, et al. (2021). Effectiveness of Expanded Food and Nutrition Education Program in Changing Nutrition-Related Outcomes among Adults with Low Income: A Systematic Review. Journal of Nutrition Education and Behavior, 53(8):691-705.

Ajie, W. N., & Chapman-Novakofski, K. M. (2014). Impact of Computer-Mediated, Obesity-Related Nutrition Education Interventions for Adolescents: A Systematic Review. J Adolesc Health, 54(6): 631-645.

Almond, D. and Currie, J. (2011). Killing Me Softly: the Fetal Origins Hypothesis. Journal of Economic Perspectives, 25(3): 153–172.

Andrabi, T., Daniels, B., & Das, J. (2021). Human Capital Accumulation and Disasters: Evidence from the Pakistan Earthquake of 2005. The Journal of Human Resources, 520.

Balbir, J. (1991). Return to education: further analysis of cross country data. Economics of education review, 1991(4):25-30.

Banerjee, A., Barnhardt, S., & Duflo, E. (2018). Can Iron-Fortified Salt Control Anemia? Evidence from Two Experiments in Rural Bihar. Journal of Development Economics, 133:127-146.

Berry, J., Mehta, S., Mukherjee, P., Ruebeck, H., Shastry, G.K. (2017). Inputs, Monitoring, and Crowd-out in India's School-Based Health Interventions. Working paper (2017).

Black, M. M. (2003). Micronutrient Deficiencies and Cognitive Functioning. The Journal of Nutrition, 133(11).

Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-preschool programs. Early Childhood Research Quarterly, 25(2), 166–176.

Buttenheim, A., Alderman, H., & Friedman, J. (2011). Impact Evaluation of School Feeding Programs in Lao PDR. Journal of Development Effectiveness, 3 (4): 520–42.

Campbell, A.A., Thorne-Lyman, A., Sun, K., De Pee, S., Kraemer, K., Moench-Pfanner, R., Sari, M., Akhter, N., Bloem, M.W. & Semba, R.D. (2008), Greater household expenditures on fruits and vegetables but not animal source foods are associated with decreased risk of under-five child mortality among families in rural Indonesia. The Journal of nutrition, 138(11): 2244-2249.

Capogrossi, K., and You, W. (2017). The Influence of School Nutrition Programs on the Weight of Low-Income Children: A Treatment Effect Analysis. Health Econ., 26: 980– 1000.

Chen, G., Tan, J., Sun, X., et al. (2017). Relationship between preschool children's neuro-psychological development and family dietary behavior. Chinese Journal of Women and Children Health, 8(05):50-54.

Cherak, S. J., Fiest, K. M., Vandersluis, L., Basualdo-Hammond, C., & Fenton, T. R. (2020). Nutrition interventions in populations with mental health conditions: a scoping review. Applied Physiology Nutrition and Metabolism, 45(2), 1-11.

de Brauw, A. and Suryanarayana, M.H. (2015). Linkages between poverty, food security and undernutrition: evidence from China and India. China Agricultural Economic Review, 7(4): 655-667.

Dhaliwal, Iqbal et. al. Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education. Education Policy in Developing Countries, edited by Paul Glewwe, University of Chicago Press, 2013.

Drewnowski, A. & Specter, S.E. (2004). Poverty and obesity: The role of energy density and energy costs, The American journal of clinical nutrition, 79(1): 6-16.

Du, Y., Kou, J., Coghill, D. (2008). The validity, reliability and normative scores of the parent, teacher and self report versions of the Strengths and Difficulties Questionnaire in China. Child Adolesc Psychiatry Ment Health, 2(1):1-15.

Frempong, R. B., & Annim, S. K. (2017). Dietary diversity and child malnutrition in Ghana. Heliyon, 3(5).

Guo, G., & Harris, K. M. (2000). The Mechanisms Mediating the Effects of Poverty on Children's Intellectual Development. Demography, 37(4): 431-447.

Hestenes, L. L., Kintner-Duffy, V., Wang, Y. C., La Paro, K., Mims, S. U., Crosby, D., Scott-Little, C., & Cassidy, D. J. (2015). Comparisons among quality measures in child care settings: Understanding the use of multiple measures in North Carolina's QRIS and their links to social-emotional development in preschool children. Early Childhood Research Quarterly, 30: 199-214.

Kazianga, H., de Walque, D. and Alderman, H. (2009). Educational and Health Impacts of Two School Feeding Schemes. Evidence from a Randomized Trial in Burkina Faso. Policy Research Working Paper 4976, World Bank, Washington, DC.

Kennedy, G. L., Pedro, M. R., Seghieri, C., Nantel, G., & Brouwer, I. (2007). Dietary Diversity Score Is a Useful Indicator of Micronutrient Intake in Non-Breast-Feeding Filipino Children. The Journal of Nutrition, 137(2): 472-477.

Kumar, A., Thapa, G.B., Joshi, P.K. (2019). Household Food Expenditure, Dietary Diversity, and Child Nutrition in Nepal. In: Thapa G., Kumar A., Joshi P. (eds) Agricultural Transformation in Nepal. Springer, Singapore.

Larson, L.M., & Yousafzai, A.K. (2017). A meta-analysis of nutrition interventions on mental development of children under-two in low- and middle-income countries. Maternal & Child Nutrition, 13(1).

Liu, Chengfang, Louise Lu, Linxiu Zhang (*), Renfu Luo, Sean Sylvia, Alexis Medina, Scott Rozelle, Darvin Scott Smith, Yingdan Chen, and Tingjun Zhu. 2017. Effect of Deworming on Indices of Health, Cognition, and Education among Schoolchildren in Rural China: A Cluster-Randomized Controlled Trial. American Journal of Tropical Medicine and Hygiene, 96(6): 1478-1489.

Luo, R., Jia, F., Yue, A., Zhang, L., Lyu, Q., Shi, Y., Yang, M., Medina, A., Kotb, S. & Rozelle, S. (2019). Passive parenting and its Association with Early Child Development, Early child development and care, 189(10): 1709-1723.

Malik, V.S., Schulze, M.B. & Hu, F.B. (2006). Intake of sugar-sweetened beverages and weight gain: A systematic review. The American journal of clinical nutrition, 84(2): 274-288.

Mbuya, M. N. N., Stoltzfus, R. J., Majo, F., Mangwadu, G., Smith, L. E., Jones, A. D., Prendergast, A. J., Jones, A. D., Tielsch, J. M., Majo, F. D., Sauramba, V., Chidhanguro, D., Mazhanga, C., Chidamba, L., Kasaru, S., Nembaware, N., Nkiwane, S., Nyamwino, P., Simango, P., . . . SHINE Trial Team. 2019. Independent and combined effects of improved water, sanitation, and hygiene, and improved complementary feeding, on child stunting and anaemia in rural zimbabwe: A cluster-randomised trial. The Lancet Global Health, 7(1): e132-e147.

McKenzie, D. (2012). Beyond baseline and follow-up: the case for more T in experiments. J Dev Econ. 99 (2): 210–21.

McEwan, P. J. (2012). Cost-effectiveness analysis of education and health interventions in developing countries. Journal of Development Effectiveness, 4(2): 189-213.

Megally, R., & Ghoneim, H. (2020). Evaluation of health intervention: A case of preschool children in Egypt. Palgrave Communications, 6(1).

Millimet, D, Tchernis, R, Husain, M. (2010). School nutrition programs and the incidence of childhood obesity. Journal of Human Resources, 45(3): 640–654.

Monsivais, P. & Johnson, D.B. (2012). Improving nutrition in home child care: are food costs a barrier?, Public health nutrition, 15(2): 370-376.

Neumann, C., Harris, D.M. & Rogers, L.M. (2002). Contribution of animal source foods in improving diet quality and function in children in the developing world, Nutrition research (New York, N.Y.), 22(1-2): 193-220.

Hair, N. L., Hanson, J. L., Wolfe, B. L., & Pollak, S. D. (2015). Association of Child Poverty, Brain Development, and Academic Achievement. JAMA pediatr, 169(9):822-829.

Olney, D. K., Pedehombga, A., Ruel, M. T., & Dillon, A. (2015). A 2-year integrated agriculture and nutrition and health behavior change communication program targeted to women in burkina faso reduces anemia,

wasting, and diarrhea in children 3-12.9 months of age at baseline: A cluster-randomized controlled trial. The Journal of Nutrition, 145(6), 1317-1324.

Onyango, A., Koski, K., & Tucker, K. (1998). Food diversity versus breastfeeding choice in determining anthropometric status in rural Kenyan toddlers. International Journal of Epidemiology, 27(3): 484-489.

Otsuka, R., Kato, Y., Nishita, Y., Tange, C., Nakamoto, M., Tomida, M., . . . Suzuki, T. (2016). Dietary diversity and 14-year decline in higher-level functional capacity among middle-aged and elderly Japanese. Nutrition, 32(7-8): 784-789.

Perkins, S. C., Finegood, E. D., & Swain, J. E. (2013). Poverty and Language Development: roles of parenting and stress. Innovations in Clinical Neuroscience, 10(4), 10-19.

Schanzenbach, D. (2009). Do school lunches contribute to childhood obesity? Journal of Human Resources, 44(3): 684–709.

Sudfeld, C. R. et al. (2015). Linear Growth and Child Development in Low- and Middle-Income Countries: A Meta-Analysis, PEDIATRICS, 135(5): e1266–e1275.

Tilak, J. B. (2007). Post-elementary education, poverty and development in India. International journal of educational development, (21): 435-445.

UNESCO. (2020). World Inequality Database on Education (WIDE), WIDE. Available at: https://www.education-inequalities.org (Accessed: 3 August 2021)

UNICEF. (2021). COVID-19: Missing More Than a Classroom. The impact of school closures on children's nutrition. Available at: https://www.unicef-irc.org/publications/pdf/COVID-19_Missing_More_Than_a_Classroom_The_impact_of_school_closures_on_childrens_nutrition.pdf (Accessed: 3 August 2021)

WFP (2020) Global Monitoring of School Meals During COVID-19 School Closures, Available at: https://cdn.wfp.org/2020/school-feeding-map/?_ga=2.106228867.1002354940.1600785227-1569974062.1594810772 (Accessed: 3 August 2021)

WFP, FAO & UNICEF. (2020). Interim Guidance Note. Mitigating the effects of the Covid-19 pandemic on food and nutrition of school children. Rome and New York, WFP, Food and Agriculture Organization of the United Nations (FAO) and UNICEF.

Whybird, G., No Tt, Z., Savage, E., Korman, N., & Chapman, J. (2020). Promoting quality of life and recovery in adults with mental health issues using exercise and nutrition intervention. International Journal of Mental Health, (1).

Zhao, W., Yu, K., Tan, S., Zheng, Y., Zhao, A., Wang, P., & Zhang, Y. (2017). Dietary diversity scores: An indicator of micronutrient inadequacy instead of obesity for Chinese children. BMC Public Health, 17(1).

Acronyms

AII	Agricultural Information Institute
ASQ:SE	Age & Stage Questionnaire: Social-Emotional
BMI	Body Mass Index
CAAS	Chinese Academy of Agricultural Sciences
CCAP	China Center for Agricultural Policy
CDRF	China Development Research Foundation
CSP	Country Strategic Plan
DDS	Dietary Diversity Scores
DEQAS	Decentralized Evaluation Quality Assurance System
DID	Difference in differences
EQ	Evaluation Question
ERG	External Reference Group
FAO	Food and Agriculture Organization
HAZ	Height-for-Age Z score
HGSF	Home-grown School Feeding
HQ	Headquarter
IFAD	International Fund for Agricultural Development
IFND	Institute of Food and Nutrition Development
IFPRI	International Food Policy Research Institute
IRB	Institutional Review Board
MARA	Ministry of Agriculture and Rural Affairs
MoE	Ministry of Education
NGO	Non-governmental Organization
NHC	National Health Commission
OSZ	Policy and Program Division
PSM	Propensity score matching
SDG	Sustainable Development Goal
SDQ	The Strengths and Difficulties Questionnaire
ToC	Theory of Change
TOR	Terms of Reference
UN	United Nations
UNEG	United Nations Evaluation Group
UNICEF	United Nations Children's Fund
WAZ	Weight-for-Age Z score
WFP	World Food Programme
WHZ	Weight-for-Height Z score
WISC-IV	Fourth edition of the Wechsler Intelligence Scale for Children
WPPSI-IV	Fourth edition of the Wechsler Preschool and Primary Scale of Intelligence

WFP CHINA OFFICE

<mark>**[Link to the website]**</mark>

World Food Programme

Via Cesare Giulio Viola 68/70
00148 Rome, Italy
T +39 06 65131 **wfp.org**