Journal of Development Effectiveness

ISSN: 1943-9342 (Print) 1943-9407 (Online) Journal homepage: https://www.tandfonline.com/loi/rjde20

Old is not always better: evidence from five randomized experiments in rural primary schools in China

Lili Li, Fang Chang, Yaojiang Shi & Scott Rozelle

To cite this article: Lili Li, Fang Chang, Yaojiang Shi & Scott Rozelle (2019) Old is not always better: evidence from five randomized experiments in rural primary schools in China, Journal of Development Effectiveness, 11:1, 68-88, DOI: 10.1080/19439342.2019.1595086

To link to this article: https://doi.org/10.1080/19439342.2019.1595086

Published online: 31 Mar 2019.

JOURNAL OF DEVELOPMENT EFFECTIVENESS
2019, VOL. 11, NO. 1, 68–88
https://doi.org/10.1080/19439342.2019.1595086

Routledge
Taylor & Francis Group

ARTICLE

Check for updates

Old is not always better: evidence from five randomized experiments in rural primary schools in China

Lili Li[a], Fang Chang[a], Yaojiang Shi[a] and Scott Rozelle[a,b]

[a]Center for Experiments Economics of Education, Shaanxi Normal University, Xi'an, China; [b]Rural Education Action Project, Stanford University, Stanford, CA, USA

ABSTRACT

In recent years, researchers have begun to focus attention on trying to identify systematic factors that cause interventions to have different impacts in different contexts. In this paper, we seek to understand whether the age of principals at schools implementing nutrition-based interventions has an impact on program outcomes. To explore the relative effectiveness of younger and older school principals, we use data from five large-scale, nutrition-related randomized controlled trials (RCTs) involving 12,595 primary school students in 336 schools in rural China. Our results, using two age cut-offs for distinguishing young principals from old ones, indicate that improvements in the health and nutrition outcomes of students were significantly higher in schools with younger principals than in schools run by older principals (when using a cutoff of 40 years old). When using a cut-off of 45 years old, the point estimates of the impacts similarly suggest that young principals are more effective, although the results are not significantly significant. The results are similar when we look at the impact of disaggregated interventions in schools managed by young and old principals. The findings are clear that the interventions implemented by older principals are not more effective than those implemented by younger principals.

ARTICLE HISTORY
Received 23 August 2017
Accepted 7 March 2019

KEYWORDS
Project effectiveness; nutrition interventions; randomized controlled trials; age of school principal; rural China

1. Introduction

In recent years, research teams have conducted impact evaluations on specific types of interventions and have often reached markedly different results. For example, several evaluation studies on mobile health projects in Africa have not reached the same conclusions on the impacts of these programs. Some papers have reported program success and positive outcomes (Rotheram-Borus et al. 2012; Lester et al. 2010), while other research appears to have found no effect (Haberer et al. 2010). Likewise, the empirical evidence on the effectiveness of CAL (computer-assisted learning) programs in promoting educational outcomes has also been mixed. For example, both Dynarski et al. (2007) and Rouse and Krueger (2004) found no significant gains in test scores in either math or reading following the implementation of CAL programs in a sample of US schools. In contrast, Barrow, Markman, and Rouse (2008) found that CAL programs improved student math test scores in Chicago schools. There are many other examples of sets of in-the-field experiments that have differential effects across studies (Maxwell et al. 2002; Quiñones et al. 1998).

Due to interest in the differential impacts found for identical interventions, researchers have begun to examine if there are systematic factors that lead to these different results. For example,

CONTACT Fang Chang changfang4421@163.com Center for Experimental Economics in Education (CEEE), Shaanxi Normal University, West Chang'an Road No. 620, Xi'an, Chang'an District 710119, China

Bold et al. (2013) launched a series of evaluations on a single type of program (also mobile health interventions). In doing so, the research team tracked the impacts of a specific treatment that was implemented by different groups in a set of randomly assigned schools. According to Bold et al. (2013), the outcomes differed depending on who implemented the intervention. In addition, Mo et al. (Forthcoming) ran a randomized controlled trial (RCT) with three treatment arms to understand whether differences in impacts of a CAL program arose based on whether an NGO or the local government implemented the project. From this research, it was clear that the effect of the CAL intervention decreased dramatically when implemented by local government rather than by an independent NGO team.

From these findings, it appears that the implementer of a project may be an important factor in determining the effectiveness of the program. This finding is of particular interest in the context of a specific program being implemented in China's rural education system. In 2012, China's national government launched a $2.2 billion per year School Feeding Program (SFP). This program is supposed to provide nutritious school lunches to over 20 million poor rural students per year in an effort to improve their health and nutrition outcomes (Ministry of Education 2011). After nearly five years of implementation, however, there is still no consensus about whether the SFP program is working or not. In one study (Wang et al. Forthcoming), results showed that the SFP did not improve overall nutrition or health outcomes; in fact, in some regions, rates of anaemia actually rose significantly. In another study (China Development Research Foundation 2017), it was found that there were general improvements in student nutrition status. Therefore, these two studies find fundamentally different outcomes when evaluating the same program in different sets of schools. Moreover, both of these studies (Wang et al. Forthcoming; China Development Research Foundation 2017) found that there were heterogeneous program impacts: namely, the SFP was more effective is some schools than others. Unfortunately, neither of these observational studies reported whether there were different outcomes between schools with different distinguishable characteristics, and, as such, the nature of the two studies limits the ability of researchers to identify reasons for these differential impacts.

In this study, we use data from five large-scale, randomized controlled trials (RCTs) that study the effectiveness of three different types of school-based, nutrition-related projects on the nutrition/health outcomes of students (namely, haemoglobin levels and the share of students with anaemia). We seek to identify if there are factors that lead to gains in nutritional outcomes in some schools and not in others. To do so, we focus on an individual characteristic of the main local implementing partner. Specifically, we examine whether the age of the school principal contributes to differential program impacts.

Two strands of literature motivate our focus on the age of the principal as a potential determinant of program success. First, there has been discussion in the public policy literature in many countries (including China) about at what age government leaders should be promoted and at what age they should be mandated to retire (Civil Servant Law of the People's Republic of China 2005). One of the underlying premises of such a policy is that younger individuals are generally thought to be more energetic and effective leaders than older individuals. If we extend this premise to the context of education, it may be the case that younger principals are more effective at implementing new programs than older principals. Unfortunately, to our knowledge, there are no empirical studies that provide causal evidence that younger managers/leaders are more effective in carrying out new policy initiatives than older ones.

Second, there are many studies in the labour economics literature that measure the relationship between age and job performance (Griffiths 1997; McEvoy and Cascio 1989; Ng and Feldman 2008). However, despite this large volume of work, empirical research findings are still mixed on the relationship between age and performance. On the one hand, there is abundant evidence that work performance improves with age (Schwab and Heneman 1977; Waldman and Avolio 1986; Giniger, Dispenzieri, and Eisenberg 1983). This likely occurs because older workers have more job experience than younger workers. This general idea has also been shown to be true in the case of teacher performance in education labour markets (Murnane 1975). Specifically, studies have found that the test scores of students in classes with older teachers who have more teaching experience are significantly higher than those of students in other classes.

On the other hand, research also provides evidence that job performance declines with age (Horn and Donaldson 1976; Craft et al. 1979; Rosen and Jerdee 1976; Skinner 1983). In a study of full-time business school faculty, Stumpf and Rabinowitz (1981) found a negative relationship between age and research productivity, speculating that intellect and mastery of important skills might decline as people age. Taylor (1975) found that older managers make slower decisions using a set of personnel decision exercises. Due in part to research findings such as these, numerous organizations have expressed concerns that older workers may exhibit lower levels of productivity (Greller and Simpson 1999; Hassell and Perrewe 1995).

Is it possible that the differences that we observe in the SFPs across schools in China are due to the age of principals? While it may be the case that this factor accounts for differential results across schools, it is unclear if the age of school principals positively or negatively impacts program success in a particular school. Additionally, considering the that current public policy literature has not evaluated this topic empirically, we have no guidance on whether we should expect younger or older school principals to be more effective administrators.

The overall goal of this study is to examine the relative effectiveness of younger and older school principals at implementing programs focused on improving the health and nutrition outcomes of students in rural China. Specifically, we seek to evaluate the differential effects of younger and older principals from schools that were involved in five large-scale RCT studies that implemented three types of nutrition-related projects in poor areas of western China. Our assumption is that if we find that the age of principals (measured with one of the two cutoffs – either 40 or 45 years old) is related to their ability to successfully implement projects that were rigorously evaluated as part of an RCT, it could help us understand whether these principals would be effective in implementing other new projects/initiatives in their school. Of course, this is an assumption, since in this study we only examine projects that focus on improving nutrition and reducing nutrition-related health problems. At the very least (even if there are no lessons outside of nutrition interventions), these findings would be of particular interest in the context China's current efforts at running its nationwide SFP (Ministry of Education 2011).

The rest of this paper is organized as follows. The next section of the paper describes our sample selection, data, and empirical methods. The third section presents the results. The final section discusses and concludes.

2. Methods

The data in this study were collected by the Rural Education Action Program (REAP), an impact evaluation organization that focuses their work in China (for more information, visit https:/reap.fsi.stanford.edu). Among its many impact evaluation efforts, REAP has employed teams of researchers to conduct RCTs to study the effects of nutrition-related initiatives on the health outcomes of students in rural Chinese schools. In the case of the projects evaluated in this study, REAP both oversaw project implementation (jointly with school principals) and conducted the evaluations.

In this paper, we draw on pooled data from five nutrition-related surveys that were conducted in rural Chinese schools. All of the projects evaluated in this study were executed between 2009 and 2012. The locations of the projects span rural areas in four provinces in western China: Shaanxi, Gansu, Ningxia, and Qinghai. Table 1 describes the sample size and treatments evaluated in each survey. Table A1 describes the provinces, grade levels, years and studies related to each survey (Kleiman-Weiner et al. 2013; Miller et al. 2012; Sylvia 2014; Wong et al. 2014; Mo et al. 2014).

Table 1. Description the study sample.

Project	Number of schools	Number of students
Full Sample	336[a]	12,595
No. 1: Gansu Vitamin Project		
Full Sub-sample	50	899
T: Chewable Vitamin	20	370
Control	30	529
No. 2: Pay for Performance 1 Project		
Full sub-sample	57	2886
T1: Information	15	596
T2: Incentive	15	667
Control	27	1623
No. 3: Pay for Performance 2 Project		
Full sub-sample	137	6446
T1: Small incentive*Small grant	20	933
T2: Large incentive*Small grant	33	1564
T3: Small incentive*Large grant	20	934
T4: Large incentive*Large grant	32	1511
Control	32	1504
No. 4: Shaanxi Vitamin Project		
Full sub-sample	60	1579
T1: Chewable Vitamin	15	390
T2: Parental Education	15	423
Control	30	766
No. 5: Ningxia Text Project		
Full sub-sample	51[b]	785
T1: Text	-	193
T2: Text and Quiz	-	201
Control	-	391

Data source: Authors' survey
Notes: [a] The total number of schools across the five projects is 336. This number does not equal the sum of all schools included in each project (355) because 19 schools were included in two or more of the projects.
[b] The interventions in this project were delivered at the individual level. Therefore, the number of schools in each treatment group is not relevant.

2.1. *Sampling*

All five surveys used a nearly identical, six-step sampling strategy. First, we obtained a list of all counties in each sample province or prefecture. Second, we randomly selected sample counties from those that met the criteria for each study, the most important criteria being population size and GDP level. Third, using official records, we created a list of all primary schools in the sample counties. Fourth, research teams used official records and telephone calls to school principals to identify whether schools met specific criteria (for example, had students in grades 1–6, were not in urban districts, and so forth.). Fifth, we randomly selected schools from the resulting sampling frame of eligible schools. Finally, we randomly selected students in the targeted age group within each sample school for inclusion in each of the studies.

After following these six steps in the five study areas, our total sample included 13,146 students in 336 schools that are located in 43 counties in four provinces. Because all five surveys were conducted as parts of RCTs, in each of the evaluation efforts the research teams assigned schools to be part of either the treatment group or control group after conducting the baseline survey. Importantly, during baseline surveying both the survey teams and the study participants were blind to their ultimate assignment into either the treatment or the control group. The number of the schools and students that were assigned into a treatment or control group was based on the power calculations which allowed the studies to be able to detect minimum effect sizes with a probability of more than 80%. When the datasets are pooled together (for this study), the power of the evaluations is significantly higher (above 90% power for identifying heterogeneous effects on two outcomes).

In summary, our study should be considered one that is using a pooled sample. This pool is made up of five independent randomized controlled trials that were all conducted and evaluated by the same research group to address the same research question. By design, the research team used different interventions to try to solve the identified problem (high anemia rates in rural schoolchildren), however, the purpose (outcome variables) of the interventions were all the same. Because of this, pooling the five datasets into one large dataset is a valid exercise and gives us an opportunity to study the question examined in this paper.

2.2. *Interventions*

Each intervention used in this study can be classified under one of three broad categories. In each of the RCTs, principals in the treatment groups were asked to implement: *treatment category 1.)* a Vitamin-based treatment; *treatment category 2.)* an Incentive-based treatment (that seeks to encourage principals to improve nutrition in the school); or *treatment category 3.)* an Information-based treatment (that also is focused on school-based nutrition problems). Table 1 shows the correspondence between the five RCTs and the different types of interventions. In three of the RCTs there was only one type of treatment (*No. 1: Gansu Vitamin Project; No. 3: Pay for Performance 2 Project; No. 5: Ningxia Text Project*), and in two of the RCTs there were two types of treatments (*No. 2: Pay for Performance 1 Project; No. 4: Shaanxi Vitamin Project*).

The first type of treatment was a Vitamin-based treatment,[1] which was conducted as follows. The vitamins that were distributed to the children were iron supplement tablets containing 5 milligrams of iron and 20 other vitamins and minerals per tablet. We trained homeroom teachers using a set intervention protocol. We asked the teachers to distribute the iron supplement tablets during their homeroom class period. They were required to have prepared a large kettle of boiled (sanitized) drinking water in advance. During the class period, the teachers would personally hand out an iron supplement tablet in one (dry) paper cup and drinking water in another cup to each child. The teachers would then monitor all children taking the tablet. Every Friday afternoon, the teachers would also give the children two extra tablets for the coming weekend and ask the children to take one on Saturday and one on Sunday. Iron supplement tablets were dispensed for six months. Periodically, the research team organized surprise inspection visits to check on levels of compliance. According to the team's observations, the levels of compliance was nearly perfect.

The second type of intervention was an Incentive-based treatment.[2] In the most general terms, the incentives were cash payments that were offered to principals in return for successfully raising haemoglobin levels and reducing anaemia rates. In each of the different individual trials the basic intervention was as follows. School principals were first provided training that focused on the prevalence of anaemia across rural China and the negative impacts of anaemia on students. Principals were also taught that there was no way to identify if an individual student had anaemia from outward symptoms. Importantly, trainers presented to the principals a number of ways to help alleviate student anaemia. After informing the principal of the actual level of anaemia in his/her school, incentive payments were offered as salary supplements to principals on a per student basis (for example, 125 yuan per student who changed from anaemic to non-anaemic over the course of the intervention). This amount was chosen since for the average principal it was equal to roughly two months of salary for successfully reducing the total number of students with anaemia by 50% – a feasible reduction according to our early pilot experience.

The third (and final) type of treatment was an Information-based treatment. This intervention, in the most general sense, was based on the idea that if parents knew that their children were anaemic, they would then take actions to reduce the incidence of anaemia. In order to make parents aware of the consequences of anaemia, the information interventions trained parents through a number of different venues and mediums (parent meetings, text messaging programs, and so forth.) about the problem of micronutrient deficiencies and their adverse impacts on children. The training materials were produced by nutritionists from the School of Medicine at

Xi'an Jiaotong University. In each of the interventions, the following information was delivered: (a) a basic description of anaemia; (b) a warning that anaemia is prevalent among children in rural areas; (c) a detailed discussion of how anaemia can affect children's health, behaviour, and learning; (d) a list of symptoms associated with anaemia; and (e) instructions on how to treat anaemia, focusing on the importance of a balanced diet that includes food with high iron content (such as red meat). The materials in each of the information interventions were simple and delivered to parents three times in three different formats: a PowerPoint presentation, a video, and a colourful booklet for each parent to take home with.

What role did the principals play in making the interventions successful or not? In any organization, including schools, the role of the head of the organization will influence the outcomes of any activity – either directly or indirectly. In all of the projects principals were instrumental in setting up the intervention – from early planning to implementation to follow up. In the vitamin project, for example, principals were charged with ensuring the teachers had sufficient time and materials (e.g. paper cups, drinking water and multiple vitamins) to carry out their work. In the incentive projects, principals were trained on the different actions that could be taken to overcome anaemia (from passing out vitamins to working with parents, etc.). In the information projects, principals were the ones that either delivered the message (training) or were present at all training sessions, showing their support. In short, principals were involved fully from the start to the end of the interventions. Therefore, the level of engagement of principals could be an important factor in the ability of an intervention to create an impact on students.

Readers that are interested in learning more details regarding each intervention can find further information though the following link: https://goo.gl/aJ4PRi.

2.3. *Ethical approval*

All five studies received ethical approval from the Stanford University Institutional Review Board (IRB). All necessary permissions also were obtained from the Chinese government as well. Oral assent was obtained from all participating children as well as the school principals, who serve as the children's legal guardians while they are in school. Principals in control schools were only told that their students were invited to participate in a study of child nutrition and that the research team would be administering a test of nutritional status. Principals in intervention schools were additionally told that their students would get the treatment (such as vitamin supplement or information for their parents) or they would get the treatment (such as their incentive package). All study participants were cognizant of the minimal risks involved and understood that their participation was purely voluntary.

2.4. *Data collection*

Our dataset can be considered a pooled dataset with different waves of observations of rural students. The surveys included in this study all followed a uniform data collection protocol. The enumerators in each survey team were local undergraduate and graduate students recruited from academic departments relevant to the content of the survey (medicine, economics, or education). All enumerators underwent comprehensive, multi-day training that lasted 2–7 days, depending on the complexity of the survey.

For each study, the research group conducted two rounds of surveys in the 336 treatment and control schools. The first round was the baseline survey and this survey was completed before any treatment implementation had begun. The second round of the survey was the evaluation survey, which was conducted at the conclusion of the experiment.

Both the baseline and evaluation surveys were nearly identical across all RCTs, and each was administered in three blocks. In the first block, the sample students were tested for their blood haemoglobin (or Hb) concentrations. Trained nurses from the Xi'an Jiaotong University School of

Medicine assessed Hb concentrations onsite using the HemoCue Hb 201+ system. These portable instruments are known to provide rapid, in-the-field measurements of Hb concentrations with high degrees of accuracy. In each survey, enumerators randomly retested 10% of all sampled students as well. If the second measure differed from the original one by more than 3.0 g/L among three or more students, the nurses retested all sample students in that school.

In the second block of the survey, enumerators collected data on the demographic and socio-economic characteristics of sample students and their families. Demographic information was collected from survey forms that were sent home by teachers and completed by the parents of sample students. Since all surveys were identical, we are able to generate control variables for student characteristics (such as gender, age, and boarding status), as well as parental characteristics (such as parental education levels and migration status). The literature provides substantial evidence these variables are important for analysing health status (Grossman 2000).

In the third block of the surveys, we also asked each principal about his/her age. We define a principal as 'young' or 'old' based on the age distribution of the principals in our sample. In the paper we use two age cut-offs, 40 and 45 years old. Therefore, in implementing our study, if the principal was less than 40 (45) years old, he/she was defined as a 'young principal.' If the principal was 40 (45) years old or older, he/she was defined as an 'old principal.' According to our data, 50% (32%) of sample principals were 40 (45) years old or older; and 50% (68%) of sample principals were younger than 40 (45) years old.[3]

In addition, in three of the RCTs, the third block of the survey collected additional school-level data from the principal of each school. We asked the principal of each school questions designed to collect information on several school characteristics, including the number of students in the school, the student-teacher ratio, whether the school had a cafeteria, the distance to the furthest village served by the schools, and the percent of students that boarded at school.

For the purpose of this study, we focused on two primary outcome measures. The first outcome measure was the altitude-adjusted Hb concentration of each student measured in g/L. Our second primary outcome measure *(also altitude-adjusted)* was the anaemia status of each student using a Hb cut-off of 120 g/L. We used anaemia as an indicator of micronutrient deficiencies in general. These two outcomes are appropriate in our study because they are the two health outcomes that the national government hoped to address when implementing the SFP (Ministry of Education 2011). Although the WHO recommends using age-specific anaemia cut-offs of 115 g/L for children aged 5–11 years and 120 g/L for those aged 12–14 years, we decided to use a 120 g/L cut-off for all children for three reasons. First, a significant proportion of our sample fell into the 12–14 years age range (25 percent of the students were older than 11 years old). Second, several studies on anaemia in rural China with sample age-distributions similar to that in our study have used the 120 g/L as an anaemia cut-off (Luo et al. 2011, 2012a). Third, in a recent paper (Sylvia and Rozelle Forthcoming), it was shown that increases in Hb concentrations that occur when a child has Hb levels in the range immediately above the WHO anaemia cut-offs (for example, in the 115 to 120 range for children 11 and younger) are associated with improvements in cognitive functioning. Note, when we redefine anaemia rates and use the age-appropriate WHO cut-offs, the fundamental findings of the study do not change.[4]

Table 3 depicts the means and standard deviations of the characteristics of the students in our combined dataset (column 1). The descriptive statistics suggest that the children are generally representative of children from rural China. For example, the gender ratio of our sample is 105 (meaning 51% of sample students are male). This is similar to the gender ratio of 104.8 reported in China's 2010 national population census (National Bureau of Statistics 2010). When looking at the average parental education levels, our data show that the share of mothers with at least a junior high school education is 28%, and the share of fathers with a junior high education or higher is 47%. In addition, 54% of students in our sample have at least one parent with a junior high school education or above. According the census data, the share of families with at least one parent who has received at least a junior high school education was 54% in 2010 (National Bureau of Statistics

2010). We also find that the share of families in our sample that had at least one parent out-migrate for work (53%) is identical to the statistic found using data from the China Health and Nutrition Survey of 53% (Li et al. 2015). In short, the characteristics of the children in our sample suggest that the data are (at least in some dimensions) representative of rural China overall. Readers that are interested in more details regarding our data collection efforts are encouraged to visit the following webpage: https://goo.gl/3n5hmj.

It should be noted that we do not control for school-level variables. We understand, however, that this might be helpful since it could be that the characteristics of schools run by young and old principals may be different. Unfortunately, we do not have access to such data for all five of the sub-datasets in the mega-dataset. As it turns out, however, we collected school-level data in three of the studies regarding four characteristics (number of students in each school; number of teachers in each school; whether the school has a kitchen and distance to the farthest village in the school's catchment area). In order to increase the power of this study, we use five interventions in our mega-dataset; because we do not have this data for all five studies we cannot control for these characteristics in our subsequent analysis. However, to address the concern that schools led by young principals might be different from schools led by old principals in ways that could affect intervention effectiveness, in Table A2 (using a 45 year old age cut-off for young and old principals) we compare the school characteristics between the schools run by young and old principals for the studies that collected school-level data. As seen, the results of this table show that, at least for the school-level variables that we have, there are no statistically significant differences between individual traits and school characteristics between schools run by young/old principals.

To address concerns that young principals might be different from old principals in terms of certain individual characteristics, we examine if there are any observable differences using survey data on the individual traits of the sample principals. Specifically, in Table A2 (also using the 45 year old age cut-off) we compare four individual characteristics (age, years of teaching experience, gender and education level) of young and old principals. As seen, the results of this table demonstrate that, at least for the individual-level variables that we have, there are no statistically significant differences between individual traits and school characteristics of schools run by young or old principals.

2.5. *Attrition and balance*

Although each project managed to track every student whose contact information was valid and was still attending school in the sample counties, the total attrition rate between the baseline and the endline surveys of the pooled data was 4.2%. Attrition was mainly due to illnesses or situations where students transferred to schools outside of our sample areas. Because of this attrition, the final sample size used in this study decreased slightly to 12,595 students.

Table 2 shows tests for attrition and balance for the mega-dataset. We conducted variable-level tests to examine the characteristics of attrited students; these results are presented in column one. We find that there are some statistically significant features of attrited students in our sample, including gender, age, Hb level at baseline, father's education and whether the father migrated. We therefore control for these variables in our subsequent analysis.

Column two shows the balance of attrited students between the treatment and control groups within the mega-dataset. In each of the five interventions, attrition did not impact the balance of treatment and control groups (see https://goo.gl/osqnzp). When we examine attrition of students in the treatment and control groups across the mega-dataset, we find that there are no statistically significant differences that would impact balance.

The final sample also shows balance across the treatment and control arms. This balance is displayed in Table 3, where we compare student and household characteristics across the treatment and control groups in the final pooled dataset (columns 2 to 4). Even in the cases

Table 2. Comparison of the characteristics of attired students and students remaining in sample and of attired students in the treatment and control groups.

Variables	Sample: sample + attrition obs. Dependent variable: attrition (1 = attrited; 0 = remained) (1)	Sample: attrition obs. Dependent variable: treatment group (1 = yes; 0 = no) (2)
Student gender (1 = male; 0 = female)	0.01*	0.07
	(0.00)	(0.05)
Student age (months)	0.0008***	−0.0004
	(0.00)	(0.00)
Boarding student (1 = yes; 0 = no)	−0.01*	−0.05
	(0.00)	(0.08)
Hb level at baseline	0.0002*	0.001
	(0.00)	(0.00)
Mother completed at least junior high school (1 = yes; 0 = no)	−0.003	−0.02
	(0.00)	(0.05)
Father completed at least junior high school (1 = yes; 0 = no)	−0.01***	−0.04
	(0.03)	(0.04)
Mother migrated (1 = yes; 0 = no)	0.01	0.01
	(0.00)	(0.05)
Father migrated (1 = yes; 0 = no)	−0.01*	0.08
	(0.00)	(0.05)
County fixed effect	Yes	Yes
N	13,146	551

Data source: Authors' survey
Notes: Cluster-robust standard errors adjusted for clustering at the school level in parentheses.
*Significant at 10%; **Significant at 5%; ***Significant at 1%

Table 3. Pre-balance test between treatment and control groups.

Variables	Full sample (1)	Treatment group (2)	Control group (3)	Difference (2)-(3) (4)
Students characteristics at baseline				
Student gender (1 = male; 0 = female)	0.51	0.51	0.51	0.00
	[0.50]	[0.50]	[0.50]	(0.01)
Student age (months)	126.9	126.9	126.8	0.15
	[16.1]	[15.5]	[17.0]	(0.76)
Boarding student (1 = yes; 0 = no)	0.16	0.15	0.19	−0.04
	[0.37]	[0.36]	[0.39]	(0.02)
Hb level at baseline	127.33	127.11	127.69	−0.58
	[12.5]	[12.5]	[12.4]	(0.75)
Family characteristics at baseline				
Mother completed at least junior high school (1 = yes; 0 = no)	0.28	0.28	0.27	0.01
	[0.45]	[0.50]	[0.44]	(0.02)
Father completed at least junior high school (1 = yes; 0 = no)	0.47	0.47	0.47	−0.00
	[0.50]	[0.50]	[0.50]	(0.02)
Mother migrated (1 = yes; 0 = no)	0.28	0.29	0.27	0.02
	[0.45]	[0.45]	[0.44]	(0.02)
Father migrated (1 = yes; 0 = no)	0.44	0.44	0.45	−0.01
	[0.50]	[0.50]	[0.50]	(0.02)
N	12,595	7782	4813	12,595

Data source: Authors' survey
Notes: a) Means with standard deviations reported in brackets.
b) Cluster-robust standard errors adjusted for clustering at the school level in parentheses.
*Significant at 10%; **Significant at 5%; ***Significant at 1%

where there are differences, the absolute magnitudes of the differences between treatment and control groups are small. The exact sampling strategies and the process of randomization, as well as attrition and balance, of the individual studies are described on the following webpage: https://goo.gl/osqnzp.

2.6. *Empirical strategy*

To understand the effect of the five interventions evaluated in our pooled dataset, and to test whether an average effect can be determined, we initially pool data from the five interventions and test whether receiving *any* treatment has an effect on haemoglobin concentrations and anaemia statuses of sample students. To estimate the effect of the interventions, we use a specification called Analysis of Covariance (ANCOVA), introduced by (Lars and Stuart 1992) and further developed by (McKenzie 2012). In multiple studies, ANCOVA has been shown to estimate more accurate results than difference-in-differences for randomized control trials (Kerwin 2014; Wydick et al. 2016). Using the ANCOVA estimation can also greatly improve the power of our analysis compared to the difference-in-differences specification (McKenzie 2012). The model is as follows:

$$\Delta Y_{ijc} = \beta_o + \beta_1 Any_{ijc} + \gamma X_{ijc} + \tau_c + \varepsilon_i \quad (1)$$

where ΔY_{ijc} is the difference in the outcome variables (that is, Hb levels or anaemia status) between baseline and endline for child *i* in school *j* in county c. The variable Any_{ijc} is represented by a dummy variable. Any_{ijc} takes on a value of 1 if student *i* receives treatment from any of the five programs and 0 otherwise. In addition, X_{ijc} represents a vector of characteristics collected at the time of the baseline survey, including student gender, age, boarding status, parental education levels and migration statuses, as well as their baseline haemoglobin concentration. The variable, τ_c, is a set of county and project fixed effects. In all specifications, standard errors are clustered at the school level. Our parameter of interest is β_1, which measures the effect of receiving treatment from any of the five programs on the average haemoglobin concentration and anaemia status of sample students.

After testing for an overall pooled effect, we then test whether the set of pooled interventions have differential impacts on the health and nutrition outcomes of students based on the age of school principals. The model we use for this purpose is:

$$\Delta Y_{ijc} = \beta_o + \beta_1 Any_{ic} + \beta_2 OLD_{ic} + \beta_3 Any_{ic} * OLD_{ic} + \gamma X_{ijc} + \tau_c + \varepsilon_i \quad (2)$$

where $\mathrm{OLD}_{\mathrm{ic}}$ is a dummy indicator representing whether the intervention in a specific school is being run by an old principal (where$\mathrm{OLD}_{\mathrm{ic}}$ takes on a value of 1) or a young principal (where $\mathrm{OLD}_{\mathrm{ic}}$ takes on a value of 0). $Any_{ic}*\mathrm{OLD}_{\mathrm{ic}}$ represents the interaction between the age of the principal (young or old) and the treatment assignment. Note, that in model (2) all other variables are the same as model (1). In the subsequent analysis we report the results for the regressions when we define OLD_{ic} using both of our age cut-offs (40 years old and 45 years old).

Next, to understand the effect of any one type of intervention (Vitamins, Incentives, or Information) on student nutrition/health outcomes, we use the following equation:

$$\Delta Y_{ijc} = \beta_o + \beta_1 \mathrm{Vitamin}_{\mathrm{ic}} + \beta_2 \mathrm{Incentive}_{\mathrm{ic}} + \beta_3 \mathrm{Information}_{\mathrm{ic}} + \gamma X_{ijc} + \tau_c + \varepsilon_i \quad (3)$$

where three dummy variables ($\mathrm{Vitamin}_{\mathrm{ic}}; \mathrm{Incentive}_{\mathrm{ic}}; \mathrm{Information}_{\mathrm{ic}}$) represent the three treatment categories: Vitamin-based treatment, Incentive-based treatment, and Information-based treatment. $\mathrm{Vitamin}_{ic}$ *or*$\mathrm{Incentive}_{ic}$ or $\mathrm{Information}_{ic}$ takes on a value of 1 if student *i* received that particular treatment and takes on the value of 0 otherwise. As above, X_{ijc} represents the same vector of baseline characteristics andτ_c represents a set of county and project fixed effects. In our regressions for model 3, standard errors also are clustered at the school level.

Finally, we also test whether the three types of interventions have different effects in young or old principal schools on student's health outcomes. The model we use for this purpose is as follows:

$$\Delta Y_{ijc} \cdot = \cdot \beta_o + \beta_1 \mathrm{Vitamin}_{\mathrm{ic}} + \beta_2 \mathrm{Incentive}_{\mathrm{ic}} + \beta_3 \mathrm{Information}_{\mathrm{ic}} + \beta_4 OLD_{ic} + \beta_5 \mathrm{Vitamin}_{\mathrm{ic}} * OLD_{ic} + \beta_6 \mathrm{Incentive}_{\mathrm{ic}} * OLD_{ic} + \beta_7 \mathrm{Information}_{\mathrm{ic}} * OLD_{ic} + \gamma X_{ijc} + \tau_c + \varepsilon_i \quad (4)$$

where the variables in this model are the same as in models (1) to (3). The only new set of variables are the three interaction terms. $Vitamin_{ic}*OLD_{ic}$ is an interaction between the age of the principal (young or old) and the Vitamin-based treatment assignment; $Incentive_{ic}*OLD_{ic}$ is an interaction between the age of the principal and the Incentive-based treatment assignment; and $Information_{ic}*OLD_{ic}$ is an interaction between the age of the principal and the Information-based treatment assignment. As in the case when we pool the interventions (see Equation 2 above), in the subsequent analysis we also report the results for the regressions when we define OLD_{ic} using both of our age cut-offs (40 years old and 45 years old).

3. Results

When we pool the data from all of the experiments together – all three types, including vitamins, incentives and information – our results (when estimating model 1) show that school interventions are able to increase haemoglobin (Hb) levels and lower anaemia rates (Table 4). Holding baseline Hb levels and student/household characteristics constant and controlling for county and project fixed effects, haemoglobin rates in the treatment schools were 1.36 g/L higher than those in the control schools following the implementation of one of the three treatments (column 1, row 1). The results also demonstrate that the anaemia rates of students in treatment schools fell by 3 percentage points relatives to those of students in control schools (column 2, row 1). Given that the average baseline anaemia rate was 23 percent, this means that, on average, when schools implemented one of the three interventions, anaemia rates dropped by 14 percent. When comparing this result to those found in the literature (Zhang et al. 2013; Luo et al. 2012a), these rates, while relatively small in magnitude, are positive and significantly different than zero, and therefore demonstrate a greater program impact than the results of a number of the studies that did not find any impacts (Cogswell et al. 2003).

Also using the full pooled dataset, Table 5 shows that the effects of implementing any of the treatments differ between schools run by younger and older principals. When the interventions are rolled out in schools managed by younger principals (using the 40 year old cut-off), student Hb levels increase by 1.43 g/L, on average, and this coefficient is significant at the 1 percent level (column 1, row 1). In contrast, the sign on the interaction term (−1.15) between the treatment and the dummy variable representing old principals is negative and statistically significant at the 5% level (column 1, row 3). The effect of the intervention on the Hb levels of students in the schools of older principals is lower than that of younger principals by 1.15 g/L. The overall treatment effect of older principals (using the 40 year old age cut-off) is statistically indistinguishable from zero (column 1, rows 10 and 11).

The same pattern holds for the regression examining the impact of any treatment on rates of anaemia (column 2). Specifically, in treatment schools run by younger principals (who were

Table 4. Impact of pooled treatments on the hemoglobin (Hb) levels and anemia rates of sample students.

	Hemoglobin levels	Anemia rates
Dependent variable	(1)	(2)
Treatment group (1 = yes; 0 = no)	1.36***	−0.03***
	(0.44)	(0.01)
Controlling for baseline Hb levels/anemia rates	Yes	Yes
Controlling for child and parent characteristics	Yes	Yes
Controlling for principal's gender and education degree	Yes	Yes
County Fixed Effects	Yes	Yes
Project Fixed Effects	Yes	Yes
N	12,595	12,595

Data source: Authors' survey
Notes: Robust standard error adjusted for clustering at the school level are reported in parentheses
*Significant at 10%; **Significant at 5%; ***Significant at 1%

Table 5. Impact of pooled treatments on the hemoglobin (Hb) levels and anemia rates of sample students between schools run by younger and older principals.

Dependent variable	Hemoglobin levels (1)	Anemia rates (2)	Hemoglobin levels (3)	Anemia rates (4)
Treatment group (1 = yes; 0 = no)	1.43***	−0.04***	0.99***	−0.03***
	(0.40)	(0.01)	(0.33)	(0.01)
Age of principal dummy (1 = older principal; 0 = younger principal)	0.57	−0.001	−0.15	0.01
	(0.63)	(0.02)	(0.71)	(0.02)
Treatment * Old (1 = yes; 0 = no)	−1.15**	0.03**	−0.35	0.01
	(0.55)	(0.01)	(0.58)	(0.02)
Controlling for baseline Hb levels/anemia rates	Yes	Yes	Yes	Yes
Controlling for child and parent characteristics	Yes	Yes	Yes	Yes
Controlling for principal's gender and diploma	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes
Project Fixed Effects	Yes	Yes	Yes	Yes
Cut off	40	40	45	45
N	12,595	12,595	12,595	12,595
Treatment Effect for Old principal school	0.28	−0.01	0.64	−0.02
P-value	0.489	0.461	0.207	0.186

Data source: Authors' survey
Notes: Robust standard error adjusted for clustering at the school level are reported in parentheses
*Significant at 10%; **Significant at 5%; ***Significant at 1%

younger than 40 years old) the anaemia rates fell by 4 percentage points (significant at the 1 percent level – row 1). Moreover, as in the case of Hb levels, we again find that there was statistically different impact on the rate of anaemia in treatment schools run by older principals (column 2, row 3). The positive and significant coefficient (0.03) suggests that the impact on anaemia was less in schools run by older principals than in schools run by younger ones. The overall treatment effect in schools run by old principals was not different than zero (rows 10 and 11). In summary, then, when using a cut-off of 40 years old (and dividing the sample approximately in half – 50% below the age cutoff and 50% above the age cutoff), the results suggest that younger principals were more effective at implementing these nutrition-based interventions in their schools than older principals. It appears that certain characteristics (such as age) of the individuals implementing interventions (such as school principals) can affect program outcomes.

When we use the alternative age cut-off (45 years old), the general nature of the results is the same with one fundamental difference (Table 5, columns 3 and 4). The signs on the variables are the same. The magnitudes of the coefficients are all similar. The overall effects on Hb levels of the younger principals are positive and significant for younger principals (column 3, row 1 or rows 10 and 11) and those on anaemia rates are negative and significant (column 4, row 1 or rows 10 and 11). However, the overall effects on the Hb levels and anaemia rates of students in the schools run by older principals are statistically indifferent from zero (columns 3 and 4, rows 12 and 13). The one difference in the results using an age cut-off of 45 years old is that, although the coefficient on the interaction terms is negative in the Hb equation (column 3, row 3) and positive for in the anaemia equation (column 4, row 3), the coefficients are not statistically different than zero. This means that there is no statistically significant difference between the impact of the programs on students in schools in which the principal is over 45 years old relative to the effects on students in schools in which the principal is under 45 years old. Hence, while the results are similar when using either the 40 or 45 year old age cut-off, the insignificance of the coefficient on the interaction term in the set of analysis using the older (45 year old) age cut-off, means caution in interpreting the results as meaning that there is absolutely clear evidence that interventions work better in the schools run by younger principals.

Table 6 presents results that also use the pooled data, but estimate the program impacts of the three different types of treatments separately. According to the findings, there are positive and significant effects of Vitamin-based and Incentive-based treatments on the Hb levels of sample

Table 6. Impact of three types of treatments on hemoglobin (Hb) levels and anemia rates of sample students.

	Hemoglobin levels	Anemia rates
Dependent variable	(1)	(2)
Vitamin-based treatment (1 = yes; 0 = no)	2.92***	−0.05*
	(0.83)	(0.03)
Incentive-based treatment (1 = yes; 0 = no)	1.64**	−0.04*
	(0.84)	(0.02)
Information-based treatment (1 = yes; 0 = no)	0.07	−0.02
	(0.44)	(0.02)
Controlling for baseline Hb levels/anemia rates	Yes	Yes
Controlling for child and parent characteristics	Yes	Yes
Controlling for principal's gender and education degree	Yes	Yes
County Fixed Effects	Yes	Yes
Project Fixed Effects	Yes	Yes
N	12,595	12,595

Data source: Authors' survey
Notes: Robust standard error adjusted for clustering at the school level are reported in parentheses
*Significant at 10%; **Significant at 5%; ***Significant at 1%

students (column 1, rows 1 and 2) and negative and significant effects on anaemia rates of sample students (column 2, rows 1 and 2). In addition, the point estimates of the impact of the Vitamin-based treatments are larger (in absolute terms), on average, than those for the Incentive-based treatments. However, the results suggest that there were no impacts of the Information-based treatments on either outcome measure (columns 1 and 2, row 3). This is in accordance with other research that found that only providing information has no impact on the health and nutrition outcomes of students (Luo et al. 2012a, 2012b). Hence, although it is not an explicit objective of our paper, Table 6 shows that the efficacy of school based-nutrition programs vary based on the nature of the intervention.

The results from the analysis using model 4 are presented in Table 7. From this table, it is possible to examine the relative effectiveness of younger and older principals at implementing each of the three different types of interventions represented in our pooled dataset. According to the results (using the 40 year old age cut-off – column 1, row 1), when examining the relative effects on Hb levels of the Incentive-based treatments, while there was significant impact in the case of schools that were run by younger principals (3.03), the estimate of the overall impact of the intervention in schools run by older principals −0.06 (column 1, rows 15 and 16) is insignificantly different than 0 and the coefficient on the interaction term (−2.65) is statistically significant. The results that show the differential impact of younger (< 40 years old) and older principals (> 40 years old) in the case of Hb levels when implementing an incentive-based system are consistent with the result of the pooled analysis (using Hb levels and the 40 year old cut-offs).

The rest of the results of the analysis that looks that the three treatments separately, however, requires more caution (in the same spirit as above). While in five of the six coefficients that measure the effect of young principals in treatment schools on Hb levels and anaemia rates (columns 2 to 4, rows 1 and 2) show that treatment students do better than control students, the levels of significance of the coefficients of the treatment and older principal interaction variables (which are all insignificantly different that zero – columns 2 to 4, rows 5 to 7)) mean that the impact of the older principals does not differ from those of younger ones. As above, however, according to these results, we are confident in saying that at the very least there is no evidence that older principals are outperforming younger principals.

Additionally, according to the results, we find significant impacts on the anaemia rates of students in schools administered by younger principals, but no significant impacts in schools run by older principals. The effect of Vitamin-based treatments on anaemia rates of students in schools run by younger principals (specifically, a reduction of 7 percentage points) are significantly greater (column 2, row 1, significant at the 10 percent level) than those in schools run by older principals

Table 7. Impact of three treatments on the hemoglobin (Hb) levels and anemia rates of sample students in schools run by younger and older principals.

	Hemoglobin levels	Anemia rates	Hemoglobin levels	Anemia rates
Dependent variable	(1)	(2)	(3)	(4)
Vitamin-based treatment (1 = yes; 0 = no)	3.03*	−0.07	3.14**	−0.07*
	(1.62)	(0.04)	(1.39)	(0.04)
Incentive-based treatment (1 = yes; 0 = no)	2.59***	−0.07***	1.69**	−0.05***
	(0.76)	(0.02)	(0.65)	(0.02)
Information-based treatment (1 = yes; 0 = no)	0.59	−0.01	0.29	−0.01
	(0.74)	(0.02)	(0.59)	(0.02)
Principal age (1 = older principal; 0 = younger principal)	0.86	0.00	−0.02	0.01
	(0.68)	(0.02)	(0.75)	(0.02)
Vitamin-based treatment * Old (1 = yes; 0 = no)	−0.33	0.02	−0.31	0.03
	(1.87)	(0.05)	(1.65)	(0.05)
Incentive-based treatment * Old (1 = yes; 0 = no)	−2.65***	0.05**	−1.39	0.02
	(0.95)	(0.02)	(1.00)	(0.02)
Information-based treatment * Old (1 = yes; 0 = no)	−0.07	0.01	1.31	−0.01
	(1.08)	(0.03)	(1.20)	(0.04)
Controlling for baseline Hb level/anemia rates	Yes	Yes	Yes	Yes
Controlling for child and parent characteristics	Yes	Yes	Yes	Yes
Controlling for principal's gender and diploma	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes
Project Fixed Effects	Yes	Yes	Yes	Yes
N	12,595	12,595	12,595	12,595
cut off	40	40	45	45
Effect of Vitamin-based treatment in schools with older principals	2.71**	−0.04	2.83**	−0.04
P-value	0.010	0.228	0.010	0.256
Effect of incentive-based treatment for schools with older principals	−0.06	−0.02	0.29	−0.02
P-value	0.935	0.392	0.760	0.251
Effect of information-based treatment for schools with older principals	0.52	−0.01	1.60	−0.02
P-value	0.477	0.729	0.117	0.668

Data source: Authors' survey
Notes: Robust standard error adjusted for clustering at the school level are reported in parentheses
*Significant at 10%; **Significant at 5%; ***Significant at 1%

(insignificant, column 2, row 13). Also, the relative effects of the Incentive-based treatments on the Hb levels and anaemia rates of students in schools run by younger principals are significantly greater than those of students in schools run by older principals (columns 1 and 2, row 2 versus row 15). The coefficients on the Incentive-based treatment variables (which measure the effect of the treatment on students in the schools run by younger principals) are both significantly different from zero (1.97 in the case of Hb levels and −0.05 in the case of anaemia rates). However, we found no significant impacts on the Hb levels or anaemia rates of students in schools run by older principals. As in the case of model 3, there were no significant measureable effects of Information-based treatments in either of the two types of schools (columns 1 and 2, row 3 and row 17).[5]

4. Extension of current results: the case of the education level of the principal

Beyond age, the labor economics literature also examines the impact of education on the outcomes of interest to individuals and society. The literature has shown that education often is an important explanatory variable (Glenda et al. 2015). In almost all papers, there is a strong and significant relationship between educational attainment and individual earnings (Loyalka et al. 2013). In other strands of the literature, the effects of education are less clear. For example, in the entrepreneur literature, there are some papers that do not find an impact of education on the performance of family firms (Dyke, Fischer, and Reuber 1992). Therefore, it is unclear if the education levels of principals should be expected to affect the effectiveness of a school-level intervention.

To examine this question, we take the same general approach as we used in the analysis of the principal age (see Equation 2 above). In our analysis we measure the education of the principal by creating a dummy variable which equals one if the principal has at least a college degree and zero if not. This variable is called *treatment_education level*. According to the descriptive statistics (Table A4), 44% of principals have a college degree; 56% do not. To measure the impact of principals with a college degree and those without, the *treatment_education level* variable is added to the equation in place of the dummy indicator representing the age of the principal.

When looking that the impact of principals without college degree, we find that there is an impact on raising hemoglobin rates and lowering anemia (Table A3). When looking at principals with a college degree, we find that there is a larger impact on hemoglobin, however this impact does not carry over for anemia (Table A3). Ultimately, the difference between the two groups is minimal. Since there is virtually no difference in principals with or without college degree, the analysis suggests that age has more explanatory power than education.

5. Summary and conclusion

In this paper, we evaluate and compare the effectiveness of younger and older principals at improving the health outcomes of students when implementing nutrition-related interventions in their schools. To do so, we present the results from five nutrition-related randomized field experiments involving 12,595 primary school students in 336 public schools in 43 national poverty counties in rural China. With this data, we draw overall conclusions using our entire sample, as well as evaluate principal effectiveness along three types of school nutrition interventions: Vitamin-based treatments; Incentive-based treatments and Information-based treatments.

The purpose of the paper is to identify whether the treatments introduced to enhance the health and nutrition of students were more successful in schools run by younger or older principals. To do this, we first show if the treatments (when taken together or evaluated separately) had an effect on measures of student haemoglobin levels and rates of anaemia. According to our findings, when we evaluate all treatments together (meaning, when a school received any treatment) or we test the effects of Vitamin-based and Incentive-based treatments separately, we find that these interventions increased the Hb levels and decreased anaemia rates among sample students. However, we find that Information-based treatments had no significant impacts on student outcomes.

Most saliently for this paper, we find fairly convincing (some caution must be exercised) that the effects of interventions differ depending on whether schools are run by younger or older principals. If we evaluate our entire sample (especially when using the 40 year old age cut-off) or those the samples receiving Vitamin-based or Incentive-based treatments, we find significant impacts on the outcome variables (and mostly significant differences) when the treatments were in schools run by younger principals. Using this same cut-off, we also find that these programs had no significant impacts (or smaller impacts) in schools administered by older principals. The results are less clear when using the 45 year old age cut-off. So, if one hesitates to use the findings to definitely conclude that the results suggest that younger principals are more effective in operationalizing school-based nutrition interventions, the results are very clear that older principals are not more effective than younger ones.

The finding that younger principals are more effective than older principals contributes to and supports the literature that younger managers are more effective leaders and, if this can be shown to be more generally true (following additional research that this paper would seem to motivate) suggests that younger individuals should be given more roles in management and public policy. The results support the findings in the labor economic literature (Greller and Simpson 1999; Hassell and Perrewe 1995). They also suggest that the finding of Stumpf and Rabinowitz (1981) that younger professors are more productive than older professors may be comparable to the case of principals.

Although we are not able to identify the exact sources of the differences in principal effectiveness, the literature would suggest that there are at least two reasons. One is that younger principals may have more energy and are willing to work harder to improve the outcomes of students in their schools. The second reason may be that younger principals are more open to new ideas and more willing to implement them in their schools. In the future, it would be valuable for researchers to attempt to evaluate the source of these differences, potentially using different research methods (such as qualitative methods) that are more capable of examining such research questions.

While the Vitamin-based and Incentive-based treatments had impacts on the haemoglobin and anaemia outcomes of students in the treatment groups, the Information-based treatment was shown to have no effect on student health outcomes. This result held when we evaluated the effects of the treatment in schools with different types of principals, as we found no significant effects in schools administered by either younger or older principals. These findings support those found in previous studies that demonstrated that information-related health treatments, on their own, are rarely effective in terms of improving health outcomes (Luo et al. 2012a, 2012b).

If these results were to be replicated and similar results were found, there are clear policy implications of these findings. For example, it may be important for schools in rural China to prioritize recruiting younger principals. Additionally, it may be appropriate to impose age limitations for school principals. We still need to find out why this is happening, but, given the fact that our results are robust across such a large sample, it appears that China should encourage further research on this topic and begin to consider new education management policies.

In assessing the findings of the paper, we do want to warn the reader to exercise caution. In some of the specifications (in some of the regression exercises), there are contradictions in the results. Specifically (as we have emphasized above), the results in many of the regressions show that interventions in the schools run by young principals in treatment schools raise Hb levels and lower anaemia rates compared to young principals in control schools. The results also suggest that the overall effects of interventions in treatment schools that are run by older principals are not significantly different than the outcomes found in control schools run by older principals. However, a contradiction arises in that in some of the specification using some of the cut-offs (i.e. the 45 year old age cut-off) the coefficients on the interaction term between older principal and treatment are statistically insignificant (although the signs and the magnitudes of the point estimates are similar to the results generated with a 40 year old age cut-off is used). Since we also have not identified the exact source of why schools with older principals have less success in these interventions, there clearly is a need for future.

In addition, it should be recognized that our results are reporting average results (ITT, heterogeneous effects). However, it is certainly true that some older principals implement interventions relatively effectively, and some younger principals do so relatively ineffectively. While it is beyond the scope of this paper to examine the factors causing these relative difference, future research teams should address this issue. Such analysis should seek to identify qualifications of principals (old and young) that are controllable and changeable, so that future policies can help to reduce the impact of age that we see currently.

Notes

1. In one of the RCT program there was one additional arm that passed out one egg per day instead of a multivitamin. Because the primary health problem of students in China's rural schools was iron-deficient anaemia, and because eggs have only trace amounts of iron, in this study we dropped the egg treatment from our pooled survey – since it was neither a true nutrition intervention nor was it a control. Importantly, however, if we include this arm as either part of the pooled vitamin treatment group or as part of the control group, the pattern of negative and positive signs and magnitudes of the coefficients in our final results are nearly the same.
2. In two of the RCT programs, there were additional arms (one in each program) that give principals a subsidy (but no incentive) to use (if he/she so desired) to improve student nutrition/health. Because the subsidy-only arm did not contain any explicit incentive, we dropped it from the pooled study since this arm was really neither a true incentive intervention nor was it a control. Importantly, however, if we include these arms as either part of the

pooled vitamin treatment group or as part of the control group, the pattern of negative and positive signs and magnitudes of the coefficients in our final results are nearly the same.

3. If we use 50 years as the cut-off of young and old principals, 92% of the sample principals were below or equal 50 years and 8% of them were older than 50 years. However, using this cut-off, the main results in our final analysis are nearly the same as if we use the 45 year old cut-off. Results are available from the authors upon request.
4. When we use the alternative definition of anaemia, the pattern of negative and positive signs and magnitudes of the coefficients in our final results are nearly the same.
5. Because we did not have access to school-level variables for all five datasets, in the analysis we did not control for school-level characteristics. As it turns out, we did have such data for three of the sub-datasets. Since differences in the schools run by young and old principals might indeed affect our results, we decided to create a subsample of our mega-dataset with three (instead of five) data sets. In this dataset we add four school-level variables (number of students in each school; number of teachers in each school; whether the school has a kitchen; and distance to the farthest village in the school's catchment area.) After doing so, we then reconducted the main analysis with this somewhat smaller school-level data set. To see if there was an impact, we then reran the analysis with and without the four school-level variables. We then compare the results of the analysis with and without the school-level data. The results, in fact, show that there are no substantive differences between the results when running the analyses with and without school-level data in the regression. Results are available from the authors upon request.

Notes on contributors

Lili Li is a PhD candidate in Economics at the Center for Experimental Economics in Education, Shaanxi Normal University. She has been researching economics of education, health and human capital in rural China for over eight years and has extensive experience in the field. Her past research has focused on how to reduce anemia in rural primary schools. She is currently working with local communities to build a sustainable model that provides quality vision care to rural communities in China.

Fang Chang is an Assistant Professor at the Center for Experimental Economics in Education, Shaanxi Normal University. She received her PhD in Economics from Northwest University. Her research has focused on improving rural education in China through nutrition and education interventions as well as experiments in teacher performance pay. She is currently working on a randomized experiment that investigates the impact of independent reading on student academic performance.

Yaojiang Shi is a Professor of Economics at the Center for Experimental Economics in Education, Shaanxi Normal University and the co-director of the Rural Education Action Project (REAP). He received his PhD in Economics from Northwest University. For twenty years his research has focused on education and economics in rural China, especially poor rural areas. In 2018, Dr. Shi was awarded the China Health Policy and Management Society Inspiration Award.

Scott Rozelle holds the Helen Farnsworth Endowed Professorship at Stanford University and is a Senior Fellow and Professor in the Freeman Spogli Institute (FSI) for International Studies. Dr. Rozelle is also the co-director of the Rural Education Action Project (REAP) and is an adjunct professor in 8 Chinese universities. Dr. Rozelle's research focuses on the economics of poverty, with an emphasis on the economics of education and health. In 2008, Dr. Rozelle was awarded China's Friendship Award—the highest honor that can be endowed on a foreign citizen—by Premiere Wen Jiabao.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the 111 Project; under Grant [number B16031]; The Outstanding Ph.D. Thesis Financial Aid Program [X2015YB08].

References

Barrow, L., L. Markman, and C. E. Rouse. 2008. "Technology's edge: The educational benefits of computer-aided instruction." *14240. NBER Working Papers*. National Bureau of Economic Research. https://ideas.repec.org/p/nbr/nberwo/14240.html

Bold, T., M. Kimenyi, G. Mwabu, A. Ng'ang'a, and J. Sandefur. 2013. "Scaling up what works: Experimental evidence on external validity in Kenyan Education." *SSRN Scholarly Paper ID 2241240*. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=2241240

China Development Research Foundation. 2017. *The Report of Students' Nutrition Status are Improved in Poor Rural Areas*. Student Nutrition Improvement Symposium. http://www.cdrf.org.cn/2017ztythbjbg/4163.jhtml

Civil Servant Law of the People's Republic of China. 2005. "The central people's government of the People's Republic of China." http://www.gov.cn/flfg/2005-06/21/content_8249.htm

Cogswell, M. E., I. Parvanta, L. Ickes, R. Yip, and G. M. Brittenham. 2003. "Iron Supplementation during Pregnancy, Anemia, and Birth Weight: A Randomized Controlled Trial." *The American Journal of Clinical Nutrition* 78 (4): 773–781. doi:10.1093/ajcn/78.4.773.

Craft, J. A., S. I. Doctors, Y. M. Shkop, and T. J. Benecki. 1979. "Simulated Management Perceptions, Hiring Decisions and Age." *Aging and Work* 2 (2): 95–102. http://scholar.google.com/scholar?cluster=7740726768344329405&hl=en&oi=scholarr

Dyke, L. S., E. M. Fischer, and A. Rebecca Reuber. 1992. "An Inter-Industry Examination of the Impact of Owner Experience on Firm Performance." *Journal of Small Business Management* 30 (4): 72–87.

Dynarski, M., R. Agodini, S. Heaviside, T. Novak, N. Carey, L. Campuzano, B. Means, et al. 2007. "Effectiveness of Reading and Mathematics Software Products: Findings from the First Student Cohort." https://telearn.archives-ouvertes.fr/hal-00190019/document

Giniger, S., A. Dispenzieri, and J. Eisenberg. 1983. "Age, Experience, and Performance on Speed and Skill Jobs in an Applied Setting." *Journal of Applied Psychology* 68 (3): 469–475. doi:10.1037/0021-9010.68.3.469.

Glenda, K., S. McGrath, I.-H. Petersen, and M. Gastrow. 2015. "Higher Education and Economic Development: The Importance of Building Technological Capabilities." *International Journal of Educational Development* 43: 22–31. doi:10.1016/j.ijedudev.2015.04.011.

Greller, M. M., and P. Simpson. 1999. "In Search of Late Career: A Review of Contemporary Social Science Research Applicable to the Understanding of Late Career." *Human Resource Management Review* 9 (3): 309–347. doi:10.1016/S1053-4822(99)00023-6.

Griffiths, A. 1997. "Ageing, Health and Productivity: A Challenge for the New Millennium." *Work & Stress* 11 (3): 197–214. doi:10.1080/02678379708256835.

Grossman, M. 2000. "The Human Capital Model of the Demand for Health, Chapter 7." *Handbook of Health Economics* 1 (January): 347–408. doi:10.1016/S1574-0064(00)80166-3.

Haberer, J. E., J. Kiwanuka, D. Nansera, I. B. Wilson, and D. R. Bangsberg. 2010. "Challenges in Using Mobile Phones for Collection of Antiretroviral Therapy Adherence Data in a Resource-Limited Setting." *AIDS and Behavior* 14 (6): 1294–1301. doi:10.1007/s10461-010-9720-1.

Hassell, B. L., and P. L. Perrewe. 1995. "An Examination of Beliefs about Older Workers: Do Stereotypes Still Exist?" *Journal of Organizational Behavior* 16 (5): 457–468. doi:10.1002/job.4030160506.

Horn, J., and G. Donaldson. 1976. "On the Myth of Intellectual Decline in Adulthood." *American Psychologist* 31(10):701.

Kerwin, J. T. 2014. "The effect of HIV infection risk beliefs on risky sexual behavior: Scared straight or scared to death?" *University of Michigan job market paper*.

Kleiman-Weiner, M., R. Luo, L. Zhang, Y. Shi, A. Medina, and S. Rozelle. 2013. "Eggs versus Chewable Vitamins: Which Intervention Can Increase Nutrition and Test Scores in Rural China?" *China Economic Review* 24 (March): 165–176. doi:10.1016/j.chieco.2012.12.005.

Lars, F., and P. Stuart. 1992. "Repeated Measures in Clinical Trials Analysis Using Mean Summary Statistics and Its Implications for Design." *Statistics in Medicine* 11: 1685–1704. doi:10.1002/sim.4780111304.

Lester, R. T., P. Ritvo, E. J. Mills, A. Kariri, S. Karanja, M. H. Chung, W. Jack, et al. 2010. "Effects of A Mobile Phone Short Message Service on Antiretroviral Treatment Adherence in Kenya (Weltel Kenya1): A Randomised Trial." *Lancet (London, England)* 376 (9755): 1838–1845. doi:10.1016/S0140-6736(10)61997-6.

Li, L., R. Luo, A. Medina, and S. Rozelle. 2015. "The Prevalence of Anemia in Central and Eastern China: Evidence from the China Health and Nutrition Survey." *The Southeast Asian Journal of Tropical Medicine and Public Health* 46 (2): 306–321.

Loyalka, P., C. Liu, Y. Song, H. Yi, X. Huang, J. Wei, L. Zhang, Y. Shi, J. Chu, and S. Rozelle. 2013. "Can Information and Counseling Help Students from Poor Rural Areas Go to High School? Evidence from China." *Journal of Comparative Economics* 41: 1012–1025. doi:10.1016/j.jce.2013.06.004.

Luo, R., Y. Shi, L. Zhang, C. Liu, S. Rozelle, B. Sharbono, A. Yue, Q. Zhao, and R. Martorell. 2012a. "Nutrition and Educational Performance in Rural China's Elementary Schools: Results of a Randomized Control Trial in Shaanxi Province." *Economic Development and Cultural Change* 60 (4): 735–772. doi:10.1086/665606.

Luo, R., Y. Shi, L. Zhang, H. Zhang, G. Miller, A. Medina, and S. Rozelle. 2012b. "The Limits of Health and Nutrition Education: Evidence from Three Randomized-Controlled Trials in Rural China." *CESifo Economic Studies* 58 (2): 385–404. doi:10.1093/cesifo/ifs023.

Luo, R., X. Wang, C. Liu, L. Zhang, Y. Shi, G. Miller, S. Rozelle, E. Yu, and R. Martorell. 2011. *Alarmingly High Anemia Prevalence in Western China*. SSRN Scholarly Paper ID 2247047. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=2247047

Maxwell, C. A., E. Msuya, M. Sudi, K. J. Njunwa, I. A. Carneiro, and C. F. Curtis. 2002. "Effect of Community-Wide Use of Insecticide-Treated Nets for 3–4 Years on Malarial Morbidity in Tanzania." *Tropical Medicine & International Health* 7 (12): 1003–1008. doi:10.1046/j.1365-3156.2002.00966.x.

McEvoy, G., and W. Cascio. February 1989. "Cumulative Evidence of the Relationship between Employee Age and Job Performance." *Journal of Applied Psychology*: 11–17. doi:10.1037/0021-9010.74.1.11.

McKenzie, D. 2012. "Beyond Baseline and Follow-Up: The Case for More T in Experiments." *Journal of Development Economics* 99: 210–221. doi:10.1016/j.jdeveco.2012.01.002.

Miller, G., R. Luo, L. Zhang, S. Sylvia, Y. Shi, P. Foo, Q. Zhao, R. Martorell, A. Medina, and S. Rozelle. 2012. "Effectiveness of Provider Incentives for Anaemia Reduction in Rural China: A Cluster Randomised Trial." *BMJ* 345 (July): e4809. doi:10.1136/bmj.e4809.

Ministry of Education. 2011. "China start to launch a rural compulsory education nutrition improvement program since 2011." http://www.moe.gov.cn/jyb_xwfb/s5148/201111/t20111130_127350.html

Mo, D., Y. Bai, Y. Shi, M. Boswell, L. Zhang, and S. Rozelle. Forthcoming. "Scaling up What Works: Evidence from A Randomized Experiment of Computer Assisted Learning in Rural China." http://reap.fsi.stanford.edu/docs/working-papers

Mo, D., R. Luo, C. Liu, H. Zhang, L. Zhang, A. Medina, and S. Rozelle. 2014. "Text Messaging and Its Impacts on the Health and Education of the Poor: Evidence from a Field Experiment in Rural China." *World Development* 64 (December): 766–780. doi:10.1016/j.worlddev.2014.07.015.

Murnane, R. J. 1975. *The Impact of School Resources on the Learning of Inner City Children*. Cambridge, MA: Balinger Publishing Company. https://eric.ed.gov/?id=ED121905

National Bureau of Statistics. 2010. "Sixth national population census." http://www.stats.gov.cn/tjsj/pcsj/rkpc/d6c/t20120718_402819792.htm

Ng, T. W. H., and D. C. Feldman. 2008. "The Relationship of Age to Ten Dimensions of Job Performance." *The Journal of Applied Psychology* 93 (2): 392–423. doi:10.1037/0021-9010.93.2.392.

Quiñones, M. L., M. C. Jo Lines, M. J. Thomson, and B. M. Greenwood. 1998. "Permethrin-Treated Bed Nets Do Not Have a 'Mass-Killing Effect' on Village Populations of Anopheles Gambiae S.L. In the Gambia." *Transactions of the Royal Society of Tropical Medicine and Hygiene* 92 (4): 373–378. doi:10.1016/S0035-9203(98)91053-7.

Rosen, B., and T. H. Jerdee. 1976. "The Influence of Age Stereotypes on Managerial Decisions." *Journal of Applied Psychology*. doi:10.1037/0021-9010.61.4.428.

Rotheram-Borus, M., M. T. Jane, W. Margaret Gwegwe, S. Comulada, N. Kaufman, and M. Keim. 2012. "Diabetes Buddies: Peer Support through a Mobile Phone Buddy System." *The Diabetes Educator* 38 (3): 357–365. doi:10.1177/0145721712444617.

Rouse, C. E., and A. B. Krueger. 2004. "Putting Computerized Instruction to the Test: A Randomized Evaluation of A 'Scientifically Based' Reading Program." *Economics of Education Review* Special Issue In Honor of Lewis C. Solman 23 (4): 323–338. doi:10.1016/j.econedurev.2003.10.005.

Schwab, D. P., and H. G. Heneman. 1977. "Age and Satisfaction with Dimensions of Work." *Journal of Vocational Behavior* 10 (2): 212–220. doi:10.1016/0001-8791(77)90058-6.

Skinner, B. F. 1983. "Intellectual Self-Management in Old Age." *The American Psychologist* 38 (3): 239–244. doi:10.1037/0003-066X.38.3.239.

Stumpf, S. A., and S. Rabinowitz. 1981. "Career Stage as a Moderator of Performance Relationships with Facets of Job Satisfaction and Role Perceptions." *Journal of Vocational Behavior* 18 (2): 202–218. doi:10.1016/0001-8791(81)90008-7.

Sylvia, S., and S. Rozelle. Forthcoming. "Anemia and cognition: Are our diagnoses wrong?" http://reap.fsi.stanford.edu/docs/working-papers

Sylvia, S. Y. 2014. "Managerial Incentives in Public Service Delivery: Evidence from School-Based Nutrition Programs in Rural China." http://drum.lib.umd.edu/handle/1903/15457

Taylor, R. N. 1975. "Age and Experience as Determinants of Managerial Information Processing and Decision Making Performance." *Academy of Management Journal* 18 (1): 74–81. doi:10.2307/255626.

Waldman, D. A., and B. J. Avolio. 1986. "A Meta-Analysis of Age Differences in Job Performance." *Journal of Applied Psychology* 71 (1): 33–38. doi:10.1037/0021-9010.71.1.33.

Wang, H., Y. Shi, W. Liang, L. Jonsson, S. Kennedy, G. Breck, J. Bai, et al. Forthcoming. "Is China's 32 Billion Dollar Program to Fight Rural Undernutrition Working? A Mixed Methods Analysis." http://reap.fsi.stanford.edu/docs/working-papers

Wong, H. L., Y. Shi, R. Luo, L. Zhang, and S. Rozelle. 2014. "Improving the Health and Education of Elementary Schoolchildren in Rural China: Iron Supplementation versus Nutritional Training for Parents." *The Journal of Development Studies* 50 (4): 502–519. doi:10.1080/00220388.2013.866223.

Wydick, B., E. Katz, F. Calvo, F. Gutierrez, and B. Janet. 2016. "Shoeing the Children: The Impact of the TOMS Shoe Donation Program in Rural El Salvador." *The World Bank*. doi:10.1093/wber/lhw042.

Zhang, L., M. Kleiman-Weiner, R. Luo, Y. Shi, R. Martorell, A. Medina, and S. Rozelle. 2013. "Multiple Micronutrient Supplementation Reduces Anemia and Anxiety in Rural China's Elementary School Children." *The Journal of Nutrition* 143 (5): 640–647. doi:10.3945/jn.112.171959.

Appendix

Table A1. Description of the five large-scale RCT samples used in this study.

Project	Location	Grade	Duration	Source
No. 1: Gansu Vitamin Project	Gansu	4th	Sep 2010-June 2011	Kleiman-Weiner et al. 2013
No. 2: Pay for Performance 1 Project	Ningxia, Qinghai	4th, 5th	Sep 2009-May 2010	Miller et al. 2012
No. 3: Pay for Performance 2 Project	Shaanxi, Qinghai, and Gansu	4th, 5th	Sep 2011-May 2012	Sylvia 2014
No. 4: Shaanxi Vitamin Project	Shaanxi	4th	Sep 2009-June 2010	Wong et al. 2014
No. 5: Ningxia Text Project	Ningxia	4th	June 2011-Oct 2012	Mo et al. 2014

Data source: Authors' survey

Table A2. Individual traits and school characteristics between young and old groups.

Variables	Full sample (1)	Old (2)	Young (3)	Difference (2)-(3)
Age-related variables				
Age (year)	40.8	49.0	36.5	12.5***
	[7.32]	[2.87]	[4.87]	(0.50)
Teaching years (year)	21.06	29.49	16.20	13.29***
	[9.34]	[5.93]	[7.27]	(0.74)
Individual traits				
Gender (1 = male,0 = female)	0.98	0.98	0.98	0.00
	[0.14]	[0.15]	[0.13]	(0.02)
Education degree (1 = with college degree,0 = without college degree)	4.00	3.91	4.05	−0.86
	[1.34]	[0.95]	[1.51]	(0.15)
School characteristics				
Student number in school	305.1	285	315	−30.5
	[256]	[238]	[264]	(47.0)
Teacher number in school	17.6	19.4	16.7	2.67
	[11.6]	[15.3]	[0.11]	(2.40)
Time to the furthest village (min)	61.9	69.1	58.0	11.1
	[55.0]	[56.2]	[54.0]	(6.76)
Has a canteen or not(1 = yes,0 = no)	0.46	0.46	0.46	−0.00
	[0.50]	[0.50]	[0.50]	(0.07)

Data source: Authors' survey
Notes: a)Means with standard deviations reported in brackets.
b) Cluster-robust standard errors adjusted for clustering at the school level in parentheses

Table A3. Impact of pooled treatments on the hemoglobin (Hb) levels and anemia rates of sample students between schools run by principals with college degree and principals without college degree.

	Hemoglobin levels	Anemia rates
Dependent variable	(1)	(2)
Treatment group (1 = yes; 0 = no)	0.69***	−0.02**
	(0.26)	(0.01)
Principals' degree (1 = with college degree;0 = without college degree)	0.34	−0.01
	(1.19)	(0.02)
Treatment * college degree (1 = yes; 0 = no)	0.77	−0.01
	(0.91)	(0.02)
Controlling for baseline Hb levels/anemia rates	Yes	Yes
Controlling for child and parent characteristics	Yes	Yes
Controlling for principal's gender	Yes	Yes
County Fixed Effects	Yes	Yes
Project Fixed Effects	Yes	Yes
N	12,595	12,595
Treatment Effect for principals without college degree	0.69***	−0.02**
P-value	0.008	0.013
Treatment Effect for principals with college degree	1.46*	−0.03
P-value	0.093	0.118

Data source: Authors' survey
Notes: Robust standard error adjusted for clustering at the school level are reported in parentheses
*Significant at 10%; **Significant at 5%; ***Significant at 1%

Table A4. Description of personal characteristics of principals in sample schools.

Characteristics of Principals	Mean (sd)
Gender of Principal (1 = male; 0 = female)	0.98 (0.16)
Principals' education degree (1 = with college degree; 0 = without college degree)	0.44 (0.50)

Data source: Authors' survey