Economics of Education Review 37 (2013) 1–12

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Do you get what you pay for with school-based health programs? Evidence from a child nutrition experiment in rural China

Sean Sylvia [a], Renfu Luo [b], Linxiu Zhang [b,*], Yaojiang Shi [c], Alexis Medina [d], Scott Rozelle [d]

[a] *Department of Agricultural and Resource Economics, University of Maryland, Symons Hall, Rm. 2200, College Park, MD 20742, USA*
[b] *Center for Chinese Agricultural Policy, Institute for Geographical Sciences and Natural Resource Research, Chinese Academy of Sciences, No. 11A Datun Road, Chaoyang District, Beijing 100101, PR China*
[c] *School of Economics and Management, Northwest University, No. 1 Xuefu Road, Guodu Education and Technology Park, Chang-an District, Xi'an 710127, Shaanxi, PR China*
[d] *Freeman Spogli Institute for International Studies, Stanford University, Encina Hall East, 5th Floor, Stanford, CA 94305, USA*

ARTICLE INFO

Article history:
Received 7 October 2012
Received in revised form 19 February 2013
Accepted 11 July 2013

JEL classification:
C93
O15
I12
I21
M52

Keywords:
Economic development
Human capital
Teacher salaries
Resource allocation
Educational economics

ABSTRACT

This study uses a randomized controlled trial of a school-based anemia reduction program in rural China to examine how increased school emphasis on health promotion affects academic performance. Although education and health promotion are complementary functions of schools, they do compete for finite school resources. We compare the effects of a traditional program that provided only information about anemia and subsidies to an otherwise identical program that included performance incentives for school principals based on school-level anemia prevalence. By the end of the trial, exam scores among students who were anemic at baseline improved under both versions of the program, but scores among students in the incentive group who were healthy at baseline fell relative to healthy students in the control group. Results suggest that performance incentives to improve student health increase the impact of school-based programs on student health outcomes, but may also lead to reallocation of school resources.

1. Introduction

School-based interventions are believed to be among the most cost-effective approaches for delivering health and nutrition services to children in developing countries (Bundy & Guyatt, 1996; Jukes, Drake, & Bundy, 2008; Orazem, Glewwe, & Patrinos, 2008). Because developing-country school systems tend to be more developed than public health systems and schools are natural points of contact with school-aged children, school systems provide a platform from which interventions can be delivered at relatively low cost (Bundy & Guyatt, 1996; Bundy et al., 2006; Jukes et al., 2008). Since improved health can in turn improve learning, the benefits of school-based health programs also include better related outcomes, such as schooling, that can improve well-being over the life course (Gomes-Neto, Hanushek, Leite, & Frota-Bezzera,1997;

* Corresponding author. Tel.: +86 010 64889834; fax: +86 010 64856533.
E-mail addresses: SSylvia@arec.umd.edu (S. Sylvia), luorf.ccap@igsnrr.ac.cn (R. Luo), lxzhang.ccap@igsnrr.ac.cn (L. Zhang), syj8882002@yahoo.com.cn (Y. Shi), amedina5@stanford.edu (A. Medina), rozelle@stanford.edu (S. Rozelle).

0272-7757/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.econedurev.2013.07.003

Orazem et al., 2008; Zhao & Glewwe, 2012; Eide & Showalter, 2011).

Despite evidence of their effectiveness, however, weak incentives for educators to improve health may be keeping school-based health and nutrition programs from reaching their full potential. Incentives facing educators in developing countries are often weak in general (cf. World Bank, 2004; Chaudhury, Hammer, Kremer, Muralidaran, & Rogers, 2006; Banerjee & Duflo, 2006; Duflo & Hanna, 2005). Further, even motivated educators may focus on traditional responsibilities over health promotion. Although health promotion and education may be complementary functions, they compete for the attention of finite school resources. Poor educator incentives for improving health may therefore reduce the ability of school-based health interventions to improve student health outcomes through reduced compliance or diversion of resources to more traditional functions.

We conducted a randomized controlled trial (RCT) in rural primary schools in western China to test whether providing school principals with pay-for-performance (P4P) contracts tied directly to health outcomes of children in their school could increase the effectiveness of a school-based anemia reduction program. Schools in the trial were allocated to either (a) a "subsidy" group in which school principals were given information about anemia and a school subsidy to implement an anemia reduction program; (b) an otherwise identical intervention that additionally provided school principals with a pay-for-performance contract based on school-level anemia prevalence (henceforth the "health incentive" group); or (c) a pure-control group.

Pay-for-performance contracts and other forms of payment tied to results – collectively known as "results-based financing" – strengthen incentives by shifting benefits to agents whose effort contributes to gains in a desired outcome. Such contracts have long been commonplace in private sector companies (e.g., sales commissions), but are now increasingly common in public service delivery (Oxman & Fretheim, 2008). In developing countries, prominent health sector examples include NGO-contracting in Cambodia that rewards use of health services (Bloom et al., 2006; Loevinsohn & Harding, 2005) and paying health facilities based on maternal and child healthcare outputs in Rwanda (Basinga et al., 2011; Gertler & Vermeersch, 2012). High-powered incentives are also being used to motivate educators – most commonly taking the form of performance pay for teachers tied to academic achievement (Hanushek & Woessmann, 2011; Woessmann, 2011).

In this paper, we focus on the impacts of the two interventions on academic performance as measured by student scores on standardized semester-end exams in math.[1] Our working hypothesis is that the subsidy intervention, through improving student health, will in turn improve student academic performance. We also hypothesize that the addition of performance incentives will lead to even stronger effects on academic performance due to larger health gains. On the other hand, it is possible that the additional emphasis placed on student health in both groups could draw attention (resources, time and effort) away from education due to a multitasking effect (Holmstrom & Milgrom, 1991; Baker, 1992; Baker, 2002). Although health promotion and education are complementary tasks (given the close relationship between good health and academic performance), they compete for finite school resources. While this type of crowding-out of educational activities is mostly a concern in the health incentive group, it is also a possibility in the subsidy group.[2]

The questions addressed in this study about the effectiveness of performance incentives in school-based health programs – and the possibility that they crowd-out educational activities – have important implications for China and other countries working to integrate nutrition into educational policy. With the explicit goal of improving nutrition among students in rural areas, the Chinese government has announced a nationwide program to provide rural students with more nutritious school meals under the "Long-term Education Reform and Development Plan (2010–2020)" (Ministry of Education, 2012). The program will initially be implemented as a pilot in 680 counties, covering about 26 million children at an annual cost of 16 billion yuan (US$2.5 billion). The majority of these funds will be given to schools as subsidies of 3 yuan per student per day (4 yuan per student per day for disadvantaged rural boarding school students). Chinese school principals (who have a significant amount of control over school expenditures) have explicit incentives for good academic performance, but preliminary fieldwork suggests that they place little emphasis on student nutrition.[3] It is therefore unclear whether subsidies to schools alone will be sufficient to achieve meaningful nutritional gains.

Our results also provide insight into the use of performance pay in public service organizations more generally. A common feature of public organizations is that they are often charged with multiple functions or roles, often for which success is not easily measured and thus they cannot be contracted upon (Dixit, 2002). In such a setting, it is possible that the introduction of performance pay tied to a subset of these functions can refocus resources away from others. School systems are prime examples of organizations with multiple roles; but our analysis speaks to a broader range of public services.

The rest of this paper is organized as follows. In Section 2 we give background on anemia, its link to educational

[1] The main results for impacts on anemia are reported in a separate paper, Miller et al. (2012).

[2] Vermeersch & Kremer (2004), for example, find evidence that a meal program in Kenyan preschools displaced teaching time by 15% despite a cook being hired to manage meal preparation. At the community level, Olken, Onishi, and Wong (2011) find evidence that incentives for health led to reductions in the provision of educational inputs. They do not, however, find reductions in academic performance and speculate that the program led to more efficient input use.

[3] School principals in China face periodic evaluations through the cadre evaluation system (see Whiting (2004)). Although the specific structure of this evaluation varies across locations, these evaluations are often based in part on student performance. In addition, approximately 20% of the principals in our sample state that they are eligible for bonus pay tied to student exam scores.

outcomes, previous studies of efforts to reduce anemia through schools and the context in which our experiment takes place. In Section 3 we describe the experimental design, sampling and the data that we use to evaluate the interventions. In Sections 4 and 5 we discuss the estimation strategy and report the effects of the subsidy and health incentive interventions on student academic performance. We also examine these findings to see which of the two (potentially offsetting) mechanisms discussed above – improved student health or reallocation of resources – appears to have the dominant effect on academic performance. The final section concludes.

2. Anemia and education in rural Northwest China

Iron deficiency is the most common nutritional deficiency worldwide (Black, 2003). In more severe cases, iron deficiency leads to anemia, a debilitating condition estimated to affect up to half of all school-aged children in developing countries (Hall et al., 2001). A large body of literature links iron deficiency – with or without anemia – to impaired cognition and brain function (Yip, 2001). More recently, iron deficiency has also been linked to attention deficit hyperactivity disorder (Konofal et al., 2008). Likely a result of these effects on cognition and behavior, iron deficient school-aged children have also been shown to have inferior educational outcomes, including grades, attendance and attainment (Nokes, van den Bosch, & Bundy, 1998; Taras, 2005).[4]

Despite rapid economic development and rising incomes in recent years, anemia rates among school-aged children in rural China remain stubbornly high. Approximately one third of children in nationally-designated poverty counties of Northwest China between ages 8 and 12 are anemic (Luo, Wang, et al., 2011; Luo et al., 2010).

Iron deficiency anemia can, in principle, be treated through relatively easy, low-cost nutrition interventions. First, greater consumption of meat, green leafy vegetables, and other iron-rich foods (as well as fruits and vegetables containing vitamin C, which promotes iron absorption) can be encouraged. Poor households may be unable to consume iron-rich foods with regularity, however, due to their inability to afford the higher priced foods.

Second, staples such as flour and soy sauce can be fortified with iron. In contrast to the success of fortification in addressing other micronutrient deficiencies, such as in iodine and Vitamin A, evidence that fortification can similarly address iron deficiency at the population level is more limited (Uauy, Hertrampf, & Reddy 2002). Further, many households in rural Northwest China grow and consume their own food (especially wheat), so fortification is likely to be ineffective.

A third approach to overcoming anemia is the provision of micronutrient supplements (multivitamins) containing iron. To be effective, however, daily consumption over time is necessary. Consequently, compliance may be inadequate due to the need for sustained effort. In addition, multivitamins are not widely available in many rural areas of China.

One method of delivering these interventions to children is to work with parents and caregivers. In low-income settings, however, multiple barriers ranging from lack of information to market imperfections limit the ability of individuals to invest in health (for a review, see Dupas (2011)). In our context, several previous studies have shown that providing caregivers with information about anemia and what can be done to reduce the risk of children becoming anemic has little effect on child health outcomes (Luo, Shi, Zhang, Liu, et al., 2012; Luo, Shi, Zhang, Zhang, et al., 2012). While more aggressive demand-side approaches – such as conditional cash transfers – may be more effective, they are also costly to implement.

Given the difficulty and cost of addressing childhood iron deficiency through demand-side interventions, working through schools on the supply-side to reduce anemia rates may be a more viable option. In a recent randomized controlled trial conducted in a setting similar to the one described in this paper, Luo, Shi, Zhang, Liu, et al. (2012) evaluate the impact of daily multivitamin provision on anemia prevalence and student performance on a standardized math exam. They find that the multivitamin intervention increased hemoglobin concentration by more than 2 g/L (or 0.2 standard deviations), on average, translating to a significant decrease in anemia. Further, standardized exam scores of anemic students in schools receiving the multivitamin intervention increased significantly.

The Luo, Shi, Zhang, Liu, et al. (2012) and Luo, Shi, Zhang, Zhang, et al. (2012) study suggests that school-based programs can be effective in reducing anemia rates. Nevertheless, that study (and most previous trials of school-based health interventions) tested only efficacy of the health intervention provided through schools. Specifically, steps were taken to ensure high compliance: field teams delivered vitamins to schools on a monthly basis, periodically supervised the teachers that passed out the vitamins (and paid them to do so), and maintained contact with schools to ensure the vitamins were being passed out daily. This level of engagement with schools, and the fact that vitamins were provided in-kind, leaves little room for results to be influenced (possibly in a negative direction) by the behavior of those in the school system who would likely implement such a program if it were scaled up.

In this study, we allow educators in the schools more control over how they choose to address anemia. The greater flexibility could lead to improvements in program effectiveness if principals and schools find more locally-appropriate strategies to address anemia. At the same time, giving principals more choice could be less effective if compliance is reduced. In the analysis here, however, we focus on another potential effect of giving schools more control over the programs: that more control opens the possibility for the intervention to affect how the school

[4] Iron repletion can improve—and possibly reverse—the detrimental effects of anemia. Improvements in language and motor development have been observed among pre-school age children in East Africa following increased levels of iron (Stoltzfus et al., 2001). In a meta-analysis of randomized controlled trials that provided iron supplements, Sachdev, Gera, and Nestel (2005) find that iron supplements significantly improved the performance of children on tests of cognitive development, especially among children who were initially anemic.

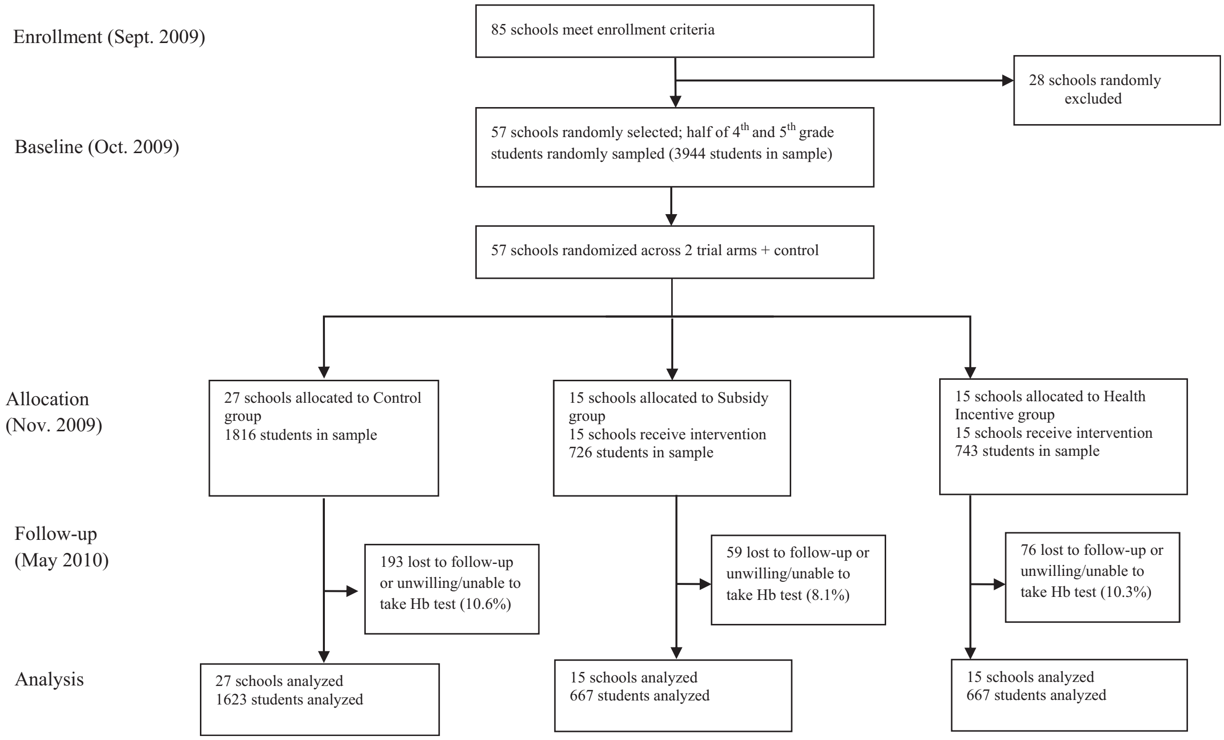

Fig. 1. Trial profile.

allocates resources and effort, in general, to other functions and across groups of students.

3. Experimental design and data collection

3.1. Sampling and randomization

To choose our sample of schools for the study, we first obtained a list of all primary schools in ten nationally-designated poverty counties in Qinghai and Ningxia.[5] In addition to limiting our focus to schools in poor counties (where anemia problems are likely to be most severe), we further limited the eligible sample to: (a) "complete" primary schools – or *wanxiao* (that is, those primary schools with the full six grades of instruction, or grades 1–6); (b) schools with at least 400 students; and (c) schools with boarding facilities. This set of criteria was chosen both to represent better the future state of rural schools given trends in the rural education system, and to reduce the possibility of significant changes at the school level that would complicate evaluation.[6] Eighty-five schools on the original list met these criteria and 57 were selected randomly for inclusion in the study.

Schools selected for inclusion in the study were then allocated randomly to a control group (27 schools) or one of two treatment groups (described below – 15 schools each). We repeated the randomization procedure until we achieved balance on pre-treatment hemoglobin concentrations with 95% confidence. The flow of schools and students through the trial are shown in Fig. 1.

3.2. Experimental design

Our experiment included a pure control group (data collection only) and two treatment groups: a *subsidy* group and a *health incentive* group. School principals in both treatment groups were provided information about anemia that focused on three points: (a) the percentage of students in their school found to be anemic at baseline[7]; (b) efficacious methods to reduce anemia (noting that school principals could implement any strategy they desire); and (c) the fact that there is scientific evidence of a correlation between anemia and impaired academic performance (according to studies conducted in China – Luo et al., 2010; Luo, Shi, Zhang, Liu, et al., 2012; Luo, Shi, Zhang, Zhang, et al., 2012). Trained enumerators conducted on-site one-on-one

[5] In China, the National Bureau of Statistics uses the designation "poverty county" to identify counties that contain significant concentrations of people living under the poverty line.

[6] Due to demographic transition and outmigration from rural areas, many areas lack a critical mass of students to justify the existence of multiple schools. Therefore, the Chinese government is consolidating many existing rural schools into new ones with these characteristics Liu, Zhang, Luo, Rozelle, & Loyalka, 2010).

[7] Principals were not told specifically which students were found to be anemic for fear of threshold effects as discussed in Neal and Whitmore Schanzenbach (2010).

training sessions with school principals lasting 1–2 h. School principals were also given pamphlets and posters about anemia, reinforcing the information provided during these health education sessions.

In addition to information, schools in the subsidy group received a subsidy of 1.5 yuan (about US$0.22) per student per day earmarked for anemia-related expenses. At the time of the baseline survey, this amount was sufficient to buy two to three ounces of red meat in local markets. The subsidy was "earmarked" in the sense that it was delivered as a grant for anemia reduction. In practice, however, we did not monitor subsidy use during the study and school principals could have used the subsidy for any purpose.[8]

Schools in the health incentive group received the same information and subsidy as the subsidy group. In addition, the school principal was given an incentive contract rewarding reductions in the number of anemic students between the baseline and endline surveys. Specifically, the contract was structured as:

$$P = \begin{cases} 150\text{RMB} * (N_b - N_e) & \text{if } (N_b - N_e) > 0 \\ 0 & \text{otherwise} \end{cases}$$

where N_b is the number of sampled students (a randomly-chosen half of all 4th and 5th graders in the school) found to be anemic at baseline and N_e is the number of these same students who were anemic at the time of the endline survey. The incremental payment of 150 yuan (US$23) per student reduction in anemia is not trivial given that the average principal salary is approximately 2500 yuan per month. We based incentive contracts on the number of sampled students for simplicity. Note that the contract includes no liability – payments to school principals were not reduced for increases in the number of anemic students.

While principals were not told the exact date of the endline survey, school principals were told the month of follow-up to give them a concrete planning horizon. Given their expectations, it is possible that principals could increase efforts to reduce anemia just before the endline survey. However, we note that it takes several months for interventions targeting anemia to be effective. This fact was stressed during school principal training sessions.

3.3. Data collection and baseline characteristics

3.3.1. Data collection

The baseline survey was conducted in September 2009 (prior to treatment assignment), and the endline survey was conducted in May 2010. Both rounds included student surveys, household surveys and school surveys. The student and household surveys collected information about socio-economic characteristics, individual health behaviors, and nutritional characteristics of meals. The school surveys collected information about school and principal characteristics.

The indicator we use for the iron status of students is altitude-adjusted hemoglobin concentration (Hb).[9] To collect hemoglobin concentration measurements, nurses from Xi'an Jiaotong Medical School accompanied enumerators during the baseline and endline surveys. Hemoglobin levels were measured on-site (at schools) using HemoCue Hb 201+ systems. This procedure is considered state-of-the-art (World Health Organization, 2001).

Our primary outcome for the analysis here is end of semester student exam scores in math *(exam scores)*.[10] Math exams are standardized at the township or county-level and scored by a panel of educators (which do not include the teachers of the students that are being tested) selected from within the county or township. Pre-treatment scores from tests at the end of the preceding school year (around June 2009) were collected by the research team during the baseline survey. Post-treatment scores during the year of the study (first semester and second semester scores) were collected during the endline survey and subsequent follow-ups with schools. First semester exams typically took place in mid-January and second semester exams typically took place in June. For analysis, we normalize scores using the control group distribution and include county-level fixed effects.

3.3.2. Sample characteristics

Table 1 summarizes baseline student hemoglobin concentration, anemia, and exam scores (Panel A) along with baseline principal characteristics (Panel B). The last three columns of the table give differences between the arms and *p*-values for the difference accounting for clustering at the school level (the unit of randomization). In the baseline survey we find that the average (altitude-adjusted) hemoglobin concentration (Hb) at baseline is around 125 g/L in all study arms and the corresponding anemia rates (defined as $Hb < 115$) are 21–25%. These figures are similar to previous studies conducted in the same area (Luo, Wang, et al., 2011; Luo, Zhang, et al., 2011b; Luo et al., 2010; Luo, Shi, Zhang, Liu, et al., 2012; Luo, Shi, Zhang, Zhang, et al., 2012). Based on the baseline characteristics in this table and additional baseline student and school characteristics in Appendix Table 1, we find no evidence of differences in observable characteristics across groups at baseline.

Panel A of Table 1 also includes raw values for student follow-up hemoglobin concentration, anemia prevalence,

[8] Given the institutional setting and the fact that others in the school know about the transfer, outright embezzlement was unlikely.

[9] At altitudes above 1000 m, it is necessary to adjust hemoglobin concentration as the distribution in normal populations increases in response to lower partial pressure of oxygen and reduced blood oxygen saturation (Nestel, 2002). To adjust measure hemoglobin, we use the following formula developed by the US CDC: $Hbl_{adj} = Hb_{measured} + 0.32 * (Altitude(m) * 0.0033) - 0.22 * (Altitude * 0.0033)^2 * 0.0033^2$.

[10] Although we also collected Chinese scores as part of the survey, we focus on math scores for two reasons. First, previous research has found that math scores are more strongly related to labor market outcomes than test scores in other subject areas (see Rose, 2006; Tyler, Murnane, & Willett, 2000). Second, compared to language scores, math scores may be more comparable across schools (particularly given that our sample includes a number of schools with significant numbers of minority students who speak Mandarin as a second language).

Table 1
Student hemoglobin concentration, anemia status, exam score, and principal characteristics.

		Control group (*C*)	Subsidy group (*T*1)	Health incentive group (*T*2)	Difference: *T*1 – C [*p*-value]	Difference: *T*2 – C [*p*-value]	Difference: *T*1 – *T*2 [*p*-value]
Panel A: Student hemoglobin concentration, anemia status, and exam scores at baseline and endline (n = 2957)							
1.	Baseline hemoglobin concentration (g/L – altitude adjusted)	125.27	123.76	124.01	−1.51	−1.26	−0.25
		(1.38)	(2.33)	(2.44)	[0.57]	[0.65]	[0.94]
2.	Endline hemoglobin concentration	129.1	129.96	130	0.86	0.9	−0.04
		(1.04)	(1.91)	(1.62)	[0.69]	[0.64]	[0.99]
3.	Anemic at baseline (0/1)	0.21	0.25	0.24	0.04	0.03	0.01
					[0.57]	[0.66]	[0.94]
4.	Anemic at endline (0/1)	0.13	0.11	0.11	−0.01	−0.02	0.01
					[0.76]	[0.56]	[0.88]
5.	Normalized baseline math score	0	−0.07	−0.3	−0.07	−0.3	0.22
		(0.08)	(0.12)	(0.19)	[0.62]	[0.16]	[0.32]
6.	Normalized Semester 1 math score	0	0.06	−0.35	0.06	−0.35	0.41
		(0.08)	(0.14)	(0.18)	[0.71]	[0.08]	[0.08]
7.	Normalized Semester 2 math score	0	−0.16	−0.09	−0.16	−0.09	−0.07
		(0.09)	(0.15)	(0.17)	[0.35]	[0.62]	[0.76]
Panel B: Principal characteristics (n = 57)							
8.	Principal age	40.63	39.27	37.8	−1.36	−2.83	1.47
		(1.04)	(1.29)	(1.92)	[0.41]	[0.2]	[0.53]
9.	Principal has college degree or above (0/1)	0.33	0.27	0.4	−0.07	0.07	−0.13
					[0.66]	[0.68]	[0.46]
10.	Principal teaching years	20.07	19.73	17.07	−0.34	−3.01	2.67
		(1.33)	(1.97)	(2.12)	[0.89]	[0.23]	[0.37]

Notes: Standard errors (in parentheses) account for clustering at the school level.

and exam scores. In the control group, mean hemoglobin increased by 3.8 g/L (likely reflecting seasonal variation in Hb levels). Larger increases occurred in the two treatment groups: mean hemoglobin concentration increased by 6.2 g/L in the Subsidy group and by 6 g/L in the Health Incentive group. Similar changes occurred for anemia rates. Relative to the control group, the mean of first semester exam scores increased by 0.13 standard deviations in the Subsidy group and fell by 0.05 standard deviations in the Health Incentive group. By the second semester, mean exam scores in the Subsidy group fell by 0.1 standard deviations relative to the control group and increased by 0.2 standard deviations in the Health Incentive group relative to the control.

3.4. Attrition and non-response

Total attrition between the baseline and endline surveys was 9.9%. Of the 3282 students surveyed during the baseline survey, 2959 were present during follow-up. In addition to those not present at follow-up, we were unable to obtain hemoglobin measurements for approximately 1% of the students that were present. Missing hemoglobin measures are not generally due to student refusal (refusal rates in both the baseline and endline were extremely low – one student in the baseline and two in the endline). Rather, they were due to test procedure error.[11]

Additionally, we were unable to obtain exam score data for approximately 20% of our sample (8% of first semester scores and 20% of second semester scores are missing). Missing test score data were largely due to students not having their grade booklets at school or exams having yet to take place at the time of data collection. To obtain grades for these students, we re-contacted schools after the endline survey. Even after multiple attempts, however, we were not able to obtain grades for a portion of our sample.

Appendix Table 2 presents analyses of how the missing data on outcome variables varied by experimental arm and baseline characteristics. The dependent variable in each model is an indicator for a missing measurement and the table reports coefficients from linear probability models.[12] The first three columns report results for missing endline hemoglobin measurements. The first column includes only the treatment arm indicators as independent variables; the second column adds other baseline covariates for the student, household and school; and the final column includes interaction terms between baseline hemoglobin concentration and the treatment arm indicators. Columns (4)–(12) repeat the same analysis for missing baseline, first semester, and second semester math exam scores.

There is no evidence that rates of missing data for primary outcomes varied across study arms. From this we can conclude that attrition is unlikely to affect the internal

[11] Usually this was difficulty obtaining a sufficient amount of blood for the test.

[12] Estimates from probit regressions are qualitatively the same. These are available upon request.

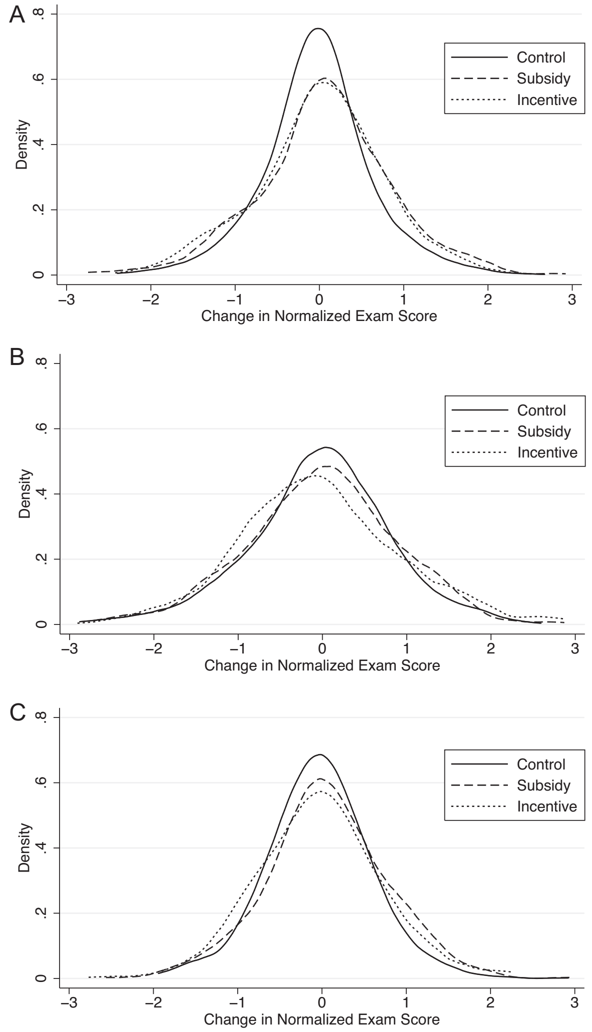

Fig. 2. Distribution of exam score changes between baseline and first second semester by treatment group. Panel A: Change between Baseline and Semester 1. Panel B: Change between Baseline and Semester 2. Panel C: Change between Baseline and Average of Semesters 1 & 2. *Notes*: Densities estimated using an Epanechnikov kernel and a bandwidth of 0.2.

validity of our estimates. Further, few of the covariates tested are significantly correlated with attrition. Only whether or not the child's mother had migrated (usually to the city for work) and student hemoglobin concentrations at baseline are consistently significantly correlated with attrition from the endline sample. While attrition correlated with baseline hemoglobin concentrations may affect the interpretation of results,[13] since this correlation does not vary across arms (columns (3), (6), (9) and (12)), the subgroup analysis by baseline hemoglobin concentrations will be unaffected by attrition.

[13] Given larger estimated effects of the interventions on students with lower baseline Hb (see results below), the negative coefficient here suggests that estimates of the main treatment effect may be lower than if there were no attrition from the sample.

4. Average program effects on standardized exam scores

Fig. 2 plots the distribution of student-level changes in exam scores by treatment group. Panel A shows the distributions of changes between baseline and first semester exams (halfway through the interventions); Panel B shows changes between baseline and second semester exams (after the conclusion of the interventions); and Panel C plots the difference between the average of students' first and second exam scores and their baseline score.[14] Most striking from these graphs is that there appears to be little difference in distribution means across the treatment groups, however both treatment group distributions seem to have a wider variance. Kolmogorov–Smirnov tests largely confirm that the two treatment distributions are indeed different than the control distribution.[15]

To quantify average treatment effects on exam scores, we estimate the model:

$$Y_{ist} = \alpha + \gamma_1 SUBSIDY_s + \gamma_2 INCENTIVE_s + \theta Y_{is(pre)} + \delta_t + X'_{is}\beta + \mu_c + \varepsilon_{is} \quad (1)$$

where Y_{ist} are normalized exam scores at follow-up; $SUBSIDY_s$ and $INCENTIVE_s$ indicate the subsidy and health incentive groups; $Y_{is(pre)}$ are pre-treatment (or baseline) exam scores from the semester before the start of the interventions; δ_t is a dummy variable for the second semester; X_{is} is a vector of other baseline characteristics included in some specifications to improve precision; μ_c is a vector of county fixed effects included to account for the fact that exams are standardized at the county or township level; and ε_{is} is an error term allowed to be correlated within schools. To reduce the effects of testing measurement error and to allow the use of all available observations, our preferred estimate of treatment effects pools first and second semester exam scores. Because it is likely that treatment effects evolve over time, however, we also present regressions with only first or second semester scores as the dependent variable. We account for missing values of baseline exam scores by setting missing values equal to zero and including a dummy variable indicating for which observations this value is missing.[16]

In this specification the parameter γ_1 identifies the effect of the subsidy intervention and γ_2 identifies the effect of the health incentive intervention. In addition to testing if γ_1 and γ_2 are significantly different from zero, we also test for equality between them (i.e., $\gamma_1 = \gamma_2$). Given our

[14] This graph uses only observations on student for whom both endline scores are available.

[15] Panel A: Subsidy-Control p-value < 0.01, Incentive-Control p-value 0.02, Subsidy-Incentive p-value 0.25; Panel B: Subsidy-Control p-value 0.14, Incentive-Control p-value 0.02, Subsidy-Incentive p-value 0.12; Panel C: Subsidy-Control p-value 0.01, Incentive-Control p-value 0.13, Subsidy-Incentive p-value 0.09.

[16] Estimates are similar if we deal with these missing values by either dropping these observations, imputing baseline exam score values using multiple imputation, or excluding baseline scores from the estimation. Results using these methods are in the appendix (Appendix Tables 4, 5, and 6).

Table 2
Average treatment effects on exam scores.

Dependent variable	Semester 1 & 2 exam scores		Semester 1 exam scores		Semester 2 exam scores	
	(1)	(2)	(3)	(4)	(5)	(6)
Subsidy group	0.05	0.05	0.13	0.13	−0.02	−0.03
	(0.085)	(0.085)	(0.093)	(0.091)	(0.121)	(0.121)
Health incentive group	−0.10	−0.11	−0.16**	−0.17**	0.00	−0.01
	(0.084)	(0.081)	(0.079)	(0.077)	(0.132)	(0.129)
Additional controls	No	Yes	No	Yes	No	Yes
Observations	5656	5656	3015	3015	2641	2641
p-Value: incentive = subsidy	0.13	0.11	0.00	0.00	0.90	0.93

Notes: Columns (1) and (2) report coefficients from specifications pooling follow-up exam scores. Columns (3)–(6) estimate semester 1 or semester 2 scores only. Exam scores are normalized by the distribution in the control group. Standard errors accounting for clustering at the school level in parentheses. All regressions control for baseline exam scores and county fixed effects. Additional controls include student sex, student age, mother's education, and migration status of the student's mother.

* $p < 0.1$.

** $p < 0.05$.

*** $p < 0.01$.

experimental design, comparing these two coefficients identifies the additive effect of providing the health incentive in addition to information and the subsidy.[17]

Table 2 reports intervention arm estimates for average effects on exam scores. Columns (1) and (2) present estimates pooling first and second semester exam scores, columns (3) and (4) present estimates for impacts on first semester scores, and columns (5) and (6) show estimates for second semester scores. Neither of the treatment effect estimates estimated with the pooled sample is distinguishable from zero. The two point estimates have opposite signs, however, and their difference is about 0.15 sd – a meaningful amount. The *p*-values in row 5 of the table suggest that this borders on statistical significance.

Moving across the table, it is clear that the largest divergence between the groups occurred during the first semester. The findings suggest that there is a negative effect of the health incentive intervention on first semester exam scores of around 0.16 sd (significant at 5%). The difference in estimated treatment effects of the two interventions is 0.3 sd (and significantly different). In the second semester, however, both estimates of the average treatment effects are close zero.

That no effect is found in the second semester could suggest that, on average, improved health did not translate into improved exam scores and that the two interventions did not lead to any indirect effects (such as a reallocation away from educational activities) that influenced exam scores. However, it is also possible that a zero effect reflects the net outcome of these two opposing forces: improved exam scores combined with resource reallocation away from educational activities. We explore this possibility further in our analysis of heterogeneous treatment effects below.

5. Heterogeneous effects by baseline hemoglobin status

It is possible that the effect of both interventions on the academic performance of students with different initial levels of hemoglobin could vary significantly. Indeed, previous trials that provided iron supplements to children suggest that – for a given amount of iron provided – anemic children experience larger gains in hemoglobin concentrations compared to non-anemic children (Luo, Shi, Zhang, Liu, et al., 2012; Luo, Shi, Zhang, Zhang, et al., 2012; Soemantri, Pollitt, & Kim, 1985; Soemantri, 1989). Moreover, the benefits of improved iron status on cognitive function and behavior may not be constant along the distribution of hemoglobin concentration (Sungthong, Mo-suwan, & Chongsuvivatwong, 2002). Children who are initially anemic, for example, may benefit more from a given improvement than initially healthy children.

To explore how the impact of the interventions on exam scores varied by baseline anemia status, we estimate heterogeneous treatment effects in five sub-groups defined by baseline hemoglobin level. Specifically, we show results for five groups: two groups in the left tail, two corresponding groups in the right tail, and a group in the middle of the distribution. The two groups in the left tail are (a) those with concentrations below 115 g/L (henceforth, *unambiguously anemic*) and (b) those with concentrations below 120 g/L (adding those who are borderline anemic to the unambiguously anemic group – henceforth *borderline anemic*). The group in the *middle* of the distribution are (c) those individuals with Hb levels that range from 115 g/L to 135 g/L. The two groups in the right tail are (d) those with concentrations above 130 g/L (*healthy group*) and (e) those with concentrations above 135 g/L (*very healthy group*). The a and b subgroups (the unambiguously and borderline anemic) represent students who are anemic and at high risk of becoming anemic at baseline and the d and e subgroups represent their healthy and very healthy counterparts.[18]

[17] To truly isolate the effect of the incentive would have required additional treatment groups and would result in a significant increase in project costs. We did not include this option, because (in addition to the additional expense and effort of running one more experimental arm), it is unlikely that a school-based anti-anemia program would not include additional resources.

[18] The distribution of baseline hemoglobin concentration in our sample is fairly symmetrically distributed around a mean of 125 g/L so this division splits observations almost evenly between lower and upper ranges.

Table 3
Nutrition intervention impacts on hemoglobin concentration and exam scores by baseline hemoglobin concentration.

		Baseline Hb < 115 g/L		Baseline Hb < 120 g/L		Baseline 115 g/L < Hb < 135 g/L		Baseline Hb > 130 g/L		Baseline Hb > 135 g/L	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Nutrition intervention impacts on hemoglobin concentration											
1.	Subsidy group	4.05***	3.84**	2.52**	2.23*	−0.26	−0.32	0.50	0.47	−0.10	−0.20
		(1.355)	(1.456)	(1.137)	(1.172)	(1.072)	(1.051)	(1.366)	(1.414)	(1.489)	(1.546)
2.	Health incentive group	3.78***	3.92***	3.55***	3.53***	1.10	1.16	0.90	1.07	1.01	1.08
		(1.317)	(1.342)	(1.134)	(1.171)	(1.053)	(1.058)	(1.197)	(1.202)	(1.586)	(1.581)
3.	Mean in control group	120.83		122.63		128.92		135.80		137.60	
4.	p-Value: incentive = subsidy	0.86	0.96	0.42	0.34	0.28	0.24	0.79	0.69	0.54	0.49
5.	Observations	718		1063		1496		998		625	
Panel B: *Nutrition intervention impacts on pooled endline exam scores*											
6.	Subsidy group	0.17**	0.18**	0.17**	0.18**	0.04	0.03	−0.02	−0.02	−0.07	−0.07
		(0.077)	(0.076)	(0.078)	(0.076)	(0.100)	(0.099)	(0.108)	(0.106)	(0.099)	(0.097)
7.	Health incentive group	0.05	0.03	0.05	0.04	−0.09	−0.10	−0.20**	−0.22***	−0.27***	−0.29***
		(0.110)	(0.108)	(0.104)	(0.102)	(0.091)	(0.089)	(0.079)	(0.077)	(0.088)	(0.090)
8.	Mean in control group	−0.08		−0.04		0.01		0.05		0.06	
9.	p-Value: incentive = subsidy	0.29	0.20	0.25	0.18	0.28	0.25	0.12	0.10	0.07*	0.05*
	Observations	1361		2030		2872		1984		1285	
10.	Additional controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Notes: Panel B reports coefficients from specifications pooling follow-up exam scores; exam scores are normalized by the distribution in the control group. Standard errors accounting for clustering at the school level in parentheses. All regressions control for baseline value of the dependent variable and county fixed effects. Additional controls include student sex, student age, mother's education, and the migration status of the student's mother.

* $p < 0.1$.

** $p < 0.05$.

*** $p < 0.01$.

Note that the two groups in the left tail (the two groups that are relatively anemic) and the two groups in the right tail (the two groups that are relatively non-anemic) are not mutually exclusive. We include estimates for both as a robustness check of our chosen cutoff values.

5.1. Hemoglobin concentration

Before moving on to exam score results, we first examine intervention effects on student hemoglobin concentrations and anemia status within each of the baseline Hb subgroups. As we discuss above, positive effects on health may not translate directly into improved exam score performance as the interventions may also influence exam scores through other channels. Examining impacts on hemoglobin concentration along the distribution can help us interpret varying impacts of the interventions on exam scores within each of the subgroups. Indeed, the relative health gains of students in the different baseline Hb subgroups provide a useful baseline for interpreting how the interventions may have affected the distribution of educational resources within schools.

Panel A of Table 3 reports estimates for effects on student endline hemoglobin concentration stratified by baseline hemoglobin concentration using an equation analogous to (1).[19] Overall, hemoglobin gains in both the subsidy and incentive arms are concentrated among those with lowest Hb levels at baseline (consistent with previous studies – e.g., Luo, Shi, Zhang, Liu, et al., 2012; Luo, Shi, Zhang, Zhang, et al., 2012; Soemantri et al., 1985; Soemantri, 1989). The increases among individuals in the unambiguously and borderline anemia groups are approximately 4 g/L and 3 g/L, respectively. None of the estimates for students in the middle, healthy or very healthy groups are significantly different from zero. On average, however, estimates in these subgroups are more positive for the incentive intervention.

5.2. Exam scores

Panel B of Table 3 presents results for impacts on exam scores by baseline hemoglobin concentration subgroups using estimations pooling first and second semester scores. For students in the unambiguously and borderline anemic groups, we find significant effects of nearly 0.2 standard deviations among students in the subsidy group (row 1; columns 1–4). These effects are on par that of with many

[19] Results for program effects on anemia status are given in the appendix (Appendix Table 3).

"successful" education interventions that have been subject to randomized evaluations (e.g., Banerjee, Cole, Duflo, & Linden 2007; Kremer, Miguel, & Thornton, 2009; Krueger & Whitmore, 2001). However, estimates for the impact of the incentive treatment among the students in the unambiguously and borderline groups are positive but not significantly different from zero. Conversely, for students in the healthy and very healthy groups, our findings suggest that the incentive intervention significantly *reduced* student exam scores by 0.2–0.3 standard deviations (row 2; columns 7–10). Wald tests show that these estimates are significantly more negative than our estimates for the effect of the subsidy intervention within the same baseline Hb subgroups (row 5). The point estimates for the effect of the subsidy intervention on exam scores of students in the healthy and very healthy subgroups are negative but not different from zero (row 1; columns 7–10).

Appendix Table 7 shows an alternative specification that, instead of estimating the effects for each subgroup separately, includes dummy variables for the two higher baseline Hb subgroup categories (the middle and very healthy groups) and interactions with treatment dummies. Although results are qualitatively identical to those in Table 4, this alternative specification allows us to formally test how treatment effects differ across the subgroups. These tests show that the effect of the subsidy treatment differs significantly between the unambiguously anemic group and the middle group (p-value = 0.02) and that the effect of the health incentive differs significantly between the middle group and the very healthy group (p-value = 0.08). The effect of the subsidy does not differ significantly between the middle group and the very healthy group (p-value = 0.17); the effect of the health incentive does not differ significantly between the unambiguously anemic group and the middle group (p-value = 0.43).

Table 4 shows estimates separately for first and second semester exam scores. Beginning with the first semester (Panel A), we find positive and significant effects of the subsidy intervention on scores of students in the unambiguously and borderline anemic groups once we include additional controls (row 1; columns 2 & 4). For the health incentive group, however, we find no statistically significant differences among anemic students. At upper ranges of the baseline hemoglobin distribution, the health incentive intervention had a significant and negative impact on exam scores of around 0.2 standard deviations (row 2; columns 7–10). Compared to the estimates for the subsidy intervention in this group, these effects are significantly more negative (row 5).

By the second semester, both interventions had large and significant effects on the exam scores of students in the unambiguously anemic subgroup: we estimate gains in test scores of around 0.3 standard deviations (rows 6 & 7; columns 1 & 2). Estimates using the subgroup of students healthy at baseline are for the large part negative (columns 5–10). Point estimates for both interventions (rows 6 & 7) are large and negative (although estimates for the subsidy group remain indistinguishable from zero). Second semester scores show even larger reductions in the health incentive group compared to the first semester with estimates ranging from −0.22 standard deviations (column 7) to −0.36 standard deviations (column 10).

We cannot say with certainty what led to these effects, but we postulate three possibilities. The first is that

Table 4
Nutrition intervention impacts on exam scores by baseline anemia status (Trend).

		Baseline Hb < 115 g/L		Baseline Hb < 120 g/L		Baseline 115 g/L < Hb < 135 g/L		Baseline Hb > 130 g/L		Baseline Hb > 135 g/L	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Semester 1 scores											
1.	Subsidy group	0.19	0.21*	0.19	0.22**	0.10	0.10	0.13	0.13	0.04	0.05
		(0.124)	(0.108)	(0.119)	(0.107)	(0.111)	(0.109)	(0.103)	(0.103)	(0.097)	(0.099)
2.	Health incentive group	−0.14	−0.17	−0.11	−0.12	−0.12	−0.12	−0.16**	−0.17**	−0.26***	−0.27***
		(0.122)	(0.113)	(0.120)	(0.111)	(0.097)	(0.096)	(0.075)	(0.078)	(0.077)	(0.079)
3.	Observations	744	744	1099	1099	1517	1517	1048	1048	681	681
4.	Mean in control group	−0.08	−0.08	−0.04	−0.04	0.01	0.01	0.05	0.05	0.06	0.06
5.	p-Value: incentive = subsidy	0.02	0.00	0.03	0.01	0.06	0.06	0.01	0.01	0.01	0.01
Panel B: Semester 2 scores											
6.	Subsidy group	0.21	0.21*	0.18	0.19	−0.03	−0.04	−0.18	−0.19	−0.25	−0.28
		(0.125)	(0.115)	(0.117)	(0.112)	(0.129)	(0.129)	(0.148)	(0.144)	(0.160)	(0.165)
7.	Health incentive group	0.29**	0.29**	0.25*	0.25*	−0.02	−0.04	−0.22*	−0.25**	−0.31*	−0.36**
		(0.138)	(0.139)	(0.127)	(0.128)	(0.139)	(0.134)	(0.118)	(0.115)	(0.154)	(0.176)
8.	Observations	617	617	931	931	1355	1355	936	936	604	604
9.	Mean in control group	−0.20	−0.20	−0.18	−0.18	0.01	0.01	0.13	0.13	0.15	0.15
10.	p-Value: incentive = subsidy	0.64	0.68	0.69	0.69	0.95	1.00	0.80	0.72	0.68	0.55
11.	Additional controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Notes: Exam scores are normalized by the distribution in the control group. Standard errors accounting for clustering at the school level in parentheses. All regressions control for baseline exam scores and county fixed effects. Additional controls include student sex, student age, mother's education, and the migration status of the student's mother.

* $p < 0.1$.

** $p < 0.05$.

*** $p < 0.01$.

educational inputs were somehow reallocated to students who were anemic at baseline. Principals or teachers could have reallocated inputs to these students consciously through targeting if they were able to ascertain which students were anemic (even though they were not given this information as part of the experiment).

Another possibility is that inputs were reallocated less advertently due to changes in behavior among initially anemic students who were pulled out of anemia by the treatment. For example, if teachers structure classes toward students who are more engaged and "ready to learn" (i.e., allocate their effort where its marginal returns are highest), the improved health of initially anemic students may lead them to demand a larger portion of teacher attention. This would explain our results, of course, with exam scores rising for the anemic students (who would be receiving relatively more teacher attention) and exam scores falling for the non-anemic students (who would be getting relatively less teacher attention).

A third possibility is that resources were allocated away from academic inputs to anemia-reduction inputs. Although all students were affected by the reduction in academic inputs, this negative effect on exam scores could have been counteracted by the positive effect of improved health for initially anemic students. Given added incentives for this type of reallocation, it is more likely to have occurred in the incentive group and may explain larger exam score reductions for initially healthy students due to this intervention.

6. Conclusion

Schools are often charged with multiple roles: increasing marketable skills of students, instilling behavioral norms, infusing national identity, and promoting health, to name a few. In developing countries, the school's role as a healthcare provider is often given added importance as school systems are used to supplement underdeveloped public health systems (Bundy et al., 2006). In this paper we take advantage of a randomized controlled trial of a school-based anemia reduction program in rural China to examine how increased emphasis on health affected student educational outcomes. Although these roles are complementary (through the relationship between health and academic performance), it is possible that the two tasks compete for limited school resources.

On average, we find that neither type of anemia reduction program – the subsidy intervention or the health incentive intervention – led to significant changes in student performance on standardized exams. We do, however, find that both versions of the anemia reduction program significantly improved the exam scores of students who were initially anemic at the start of the trial but reduced exam scores for initially healthy students. Both of these effects were more pronounced when school principals were given incentives for health improvement in the form of performance pay based on students' anemia status. One interpretation of these findings is that the added emphasis on improving student health led to a redistribution of school resources – either by a (conscious or unconscious) reallocation of inputs to students who were initially anemic, or by an overall reallocation of school resources from academic inputs to anemia reduction inputs – and that this redistribution was larger when principals were given additional incentives for health improvement.

A number of caveats should be considered along with the results presented here. First, as with any achievement test, it is unclear how closely the county-level exam scores we use reflect true learning. While we address this in a limited way by using scores from two post-treatment periods, they presumably contain significant measurement error. Still, given the weight placed on such exams in the Chinese education system, they are important in their own right. Second, we were unable to obtain exam scores for a significant portion of the sample. Given that this loss is balanced across treatment groups, we do not believe it affects the internal validity of the study; however, it is possible that the analytic sample deviates in some way from the true achievement distribution in the schools. Finally, our study takes place in poor rural regions of northwest China – results may vary in other settings, particularly those with significantly different incentive systems.

Nevertheless, we believe these results offer insights into the design of incentives in public service organizations charged with multiple functions. They also have direct implications for the implementation of school-based health and nutrition programs. Policy makers should consider carefully how adding the additional role of promoting student health – and strengthening incentives in this direction – interacts with existing, traditional school functions. Achieving the desired balance between purely educational functions and health-improvement functions requires a comprehensive consideration of the incentive structure embedded within the education system.

Acknowledgements

We are grateful to the Chinese Academy of Sciences (KZZD-EW-06-02), Institute of Geographic Sciences and Natural Resources Research, Chinese. Academy of Sciences (2011RC102), the National Natural Science Foundation of China(71033003),International Initiative for Impact Evaluation (3ie), and Stanford University (Freeman Spogli Institute for International Studies, Stanford Center for International Development, and The Walter H. Shorenstein Asia-Pacific Research Center) for financial support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.econedurev.2013.07.003.

References

Baker, G. (2002). Distortion and risk in optimal incentive contracts. *The Journal of Human Resources, 37*(4), 728–751.

Baker, G. (1992). Incentive contracts and performance measurement. *Journal of Political Economy, 100*(3), 598–614.

Banerjee, A., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *The Quarterly Journal of Economics, 122*(3), 1235–1264.

Banerjee, A., & Duflo, E. (2006). Addressing absence. *The Journal of Economic Perspectives, 20*(1), 117–132.

Basinga, P., Gertler, P., Binagwaho, A., Soucat, A., Sturdy, J., & Vermeersch, C. (2011). Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: An impact evaluation. *The Lancet, 377*(9775), 1421–1428.

Black, M. (2003). Micronutrient deficiencies and cognitive functioning. *The Journal of Nutrition, 133*(11), 3927S–3931S.

Bloom, E., King, E., Bhushan, I., Kremer, M., Clingingsmith, D., Loevinsohn, B., et al. (2006). *Contracting for health: Evidence from Cambodia*. Brookings Institution.

Bundy, D., & Guyatt, H. (1996). Schools for health: Focus on health, education and the school-age child. *Parasitology Today, 12*(8), 1–14.

Bundy, D., Shaeffer, S., Jukes, M., Beegle, K., Gillespie, A., Drake, L., et al. (2006). *School-based health and nutrition programs. In disease control priorities in developing countries, 30:75*. World Bank, World Health Organization, National Institutes of Health.

Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K., & Rogers, F. (2006). Missing in action: Teacher and health worker absence in developing countries. *The Journal of Economic Perspectives, 20*(1), 91–116.

Dixit, A. (2002). Incentives and organizations in the public sector: An interpretative review. *The Journal of Human Resources, 37*(4), 696–727.

Duflo, E., & Hanna, R. (2005). *Monitoring works: Getting teachers to come to school*. National Bureau of Economic Research.

Dupas, P. (2011). Health behavior in developing countries. *Annual Review of Economics, 3*, 425–449.

Eide, E., & Showalter, M. (2011). Estimating the relation between health and education: What do we know and what do we need to know? *Economics of Education Review, 30*(5), 778–791.

Gertler, P., & Vermeersch, C. (2012). *Using Performance Incentives to Improve Health Outcomes*. Working Paper.

Gomes-Neto, J. B., Hanushek, E., Leite, R. H., & Frota-Bezzera, R. C. (1997). Health and schooling: Evidence and policy implications for developing countries. *Economics of Education Review, 16*(3), 271–282.

Hall, A., Bobrow, E., Brooker, S., Jukes, M., Nokes, K., Lambo, J., et al. (2001). Anaemia in schoolchildren in eight countries in Africa and Asia. *Public Health Nutrition, 4*(3), 749–756.

Hanushek, E., & Woessmann, L. (2011). Overview of the symposium on performance pay for teachers. *Economics of Education Review, 30*(3), 391–393.

Holmstrom, B., & Milgrom, P. (1991). Multitask principal – agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization, 7*(special issue), 24–52.

Jukes, M., Drake, L., & Bundy, D. (2008). *School Health, Nutrition and Education for All: Levelling the Playing Field*. CABI.

Konofal, E., Lecendreux, M., Deron, J., Marchand, M., Cortese, S., Zaïm, M., et al. (2008). Effects of iron supplementation on attention deficit hyperactivity disorder in children. *Pediatric Neurology, 38*(1), 20–26.

Kremer, M., Miguel, E., & Thornton, R. (2009). Incentives to learn. *Review of Economics and Statistics, 91*(3), 437–456.

Krueger, A., & Whitmore, D. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from project STAR. *The Economic Journal, 111*(468), 1–28.

Liu, C., Zhang, L., Luo, R., Rozelle, S., & Loyalka, P. (2010). The effect of primary school mergers on academic performance of students in rural China. *International Journal of Educational Development, 30*(6), 570–585.

Loevinsohn, B., & Harding, A. (2005). Buying results? Contracting for health service delivery in developing countries. *The Lancet, 366*(9486), 676–681.

Luo, R., Kleiman-Weiner, M., Rozelle, S., Zhang, L., Liu, C., Sharbono, B., et al. (2010). Anemia in rural China's elementary schools: Prevalence and correlates in Shaanxi Province's poor counties. *Ecology of Food and Nutrition, 49*(5), 357–372.

Luo, R., Wang, X., Zhang, L., Liu, C., Shi, Y., Miller, G., et al. (2011). Alarmingly high anemia prevalence in Western China. *Southeast Asian Journal of Tropical Medicine and Public Health42.(5).*

Luo, R., Shi, Y., Zhang, L., Liu, C., Rozelle, S., Sharbono, B., et al. (2012). 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.

Luo, R., Shi, Y., Zhang, L., Zhang, H., Miller, G., Medina, A., et al. (2012). The limits of health and nutrition education: Evidence from three randomized-controlled trials in rural China. *CESifo Economic Studies, 58*(2), 385–404.

Luo, R., Zhang, L., Liu, C., Zhao, Q., Shi, Y., Miller, G., et al. (2011). Anaemia among students of rural China's elementary schools: Prevalence and correlates in Ningxia and Qinghai's Poor Counties. *Journal of Health, Population and Nutrition, 29*(5), 471–485.

Miller, G., Luo, R., Zhang, L., Sylvia, S., Shi, Y., Foo, P., et al. (2012). Effectiveness of provider incentives for anemia reduction in rural China: A cluster randomized trial. *British Medical Journal, 345*, e4809.

Ministry of Education. (2012). The Guideline of Implementing the Nutrition Improvement Program for Compulsory Education Students in Rural China.

Neal, D., & Whitmore Schanzenbach, D. (2010). Left behind by design: Proficiency counts and test-based accountability. *Review of Economics and Statistics, 92*(2), 263–283.

Nestel, P. (2002). *Adjusting hemoglobin values in program surveys*. Washington, DC: International Nutritional Anaemia Consultative Group, ILSI Human Nutrition Institute.

Nokes, C., van den Bosch, C., & Bundy, D. (1998). *The effects of Iron deficiency and anemia on mental and motor performance, educational achievement, and behavior in children. A report of the INACG*. Washington, DC: International Life Sciences Institute.

Olken, B., Onishi, J., & Wong, S. (2011). *Should aid reward performance? Evidence from a field experiment on health and education in Indonesia* Mimeo MIT.

Orazem, P., Glewwe, P., & Patrinos, H. (2008). The benefits and costs of alternative strategies to improve educational outcomes. *Copenhagen Consensus 2008 Challenge Paper: Education.*

Oxman, A., & Fretheim, A. (2008). An Overview of research on the effects of results-based financing. *Nasjonalt kunnskapssenter for helsetjenesten (Norwegian Knowledge Centre for the Health Services).*

Rose, H. (2006). Do gains in test scores explain labor market outcomes? *Economics of Education Review, 25*(4), 430–446.

Sachdev, H., Gera, T., & Nestel, P. (2005). Effect of Iron supplementation on mental and motor development in children: Systematic review of randomised controlled trials. *Public Health Nutrition, 8*(2), 117–132.

Soemantri, A. (1989). Preliminary findings on iron supplementation and learning achievement of rural Indonesian children. *The American Journal of Clinical Nutrition, 50*(3), 698–702.

Soemantri, A., Pollitt, E., & Kim, I. (1985). Iron deficiency anemia and educational achievement. *The American Journal of Clinical Nutrition, 42*(6), 1221–1228.

Stoltzfus, R., Kvalsvig, J., Chwaya, H., Montresor, A., Albonico, M., Tielsch, J., et al. (2001). Effects of iron supplementation and anthelmintic treatment on motor and language development of preschool children in zanzibar: Double blind, placebo controlled study. *BMJ (Clinical Research Ed.), 323*(7326), 1389–1393.

Sungthong, R., Mo-suwan, L., & Chongsuvivatwong, V. (2002). Effects of haemoglobin and serum ferritin on cognitive function in school children. *Asia Pacific Journal of Clinical Nutrition, 11*(2), 117–122.

Taras, H. (2005). Nutrition and student performance at school. *Journal of School Health, 75*(6), 199–213.

Tyler, J., Murnane, R., & Willett, J. (2000). Estimating the labor market signaling value of the GED. *The Quarterly Journal of Economics, 115*(2), 431–468.

Uauy, R., Hertrampf, E., & Manju, R. (2002). Iron fortification of foods: Overcoming technical and practical barriers. *The Journal of Nutrition, 132*(4), 849S–852S.

Vermeersch, C., & Kremer, M. (2004). School Meals. *Educational Achievement and School Competition: Evidence from a Randomized Evaluation* World Bank Policy Research Working Paper No. 3523.

Whiting, S. (2004). The cadre evaluation system at the grass roots: The paradox of Party rule. In B. Naughton & D. Yang (Eds.), *Holding China together: Diversity and national integration in the post-Deng era* (pp. 101–119). New York: Cambridge University Press.

Woessmann, L. (2011). Cross-country evidence on teacher performance pay. *Economics of Education Review, 30*(3), 404–418.

World Bank (2004). *World Development Report 2004: Making Services Work for Poor People*. World Bank, Washington, DC.

World Health Organization (2001). *Iron Deficiency Anemia: Assessment, Prevention, and Control – A Guide for Program Managers*. Geneva, Switzerland: World Health Organization.

Yip, R. (2001). Iron Deficiency and Anemia. In R. D. Semba & M. W. Bloem (Eds.), *Nutrition and Health in Developing Countries* (pp. 327–342). New York: Humana Press.

Zhao, M., & Glewwe, P. (2012). What determines basic school attainment in developing countries? Evidence from rural china. *Economics of Education Review, 29*(3), 451–460.