Evolution of Evaluation

Scott Rozelle Freeman Spogli Institute, Stanford University

Rural Education Action Project (REAP)

My assignment:

"In particular, we would encourage you to reflect on how impact evaluation work in development has changed in the last decade."

Some questions relevant to this topic include:

- Are more studies being conducted? [Q1]
- Is the quality improving?
- How are the methods that are utilized changing over time?/ [Q2]
- Is capacity growing in foreign aid agencies and developing countries?
- Are study findings making their way into policy decisions in developing countries and at organizations like the multilateral development banks?
- What problems might be emerging in terms of ethics, capacity, quality of evidence, and dissemination or use? Q3

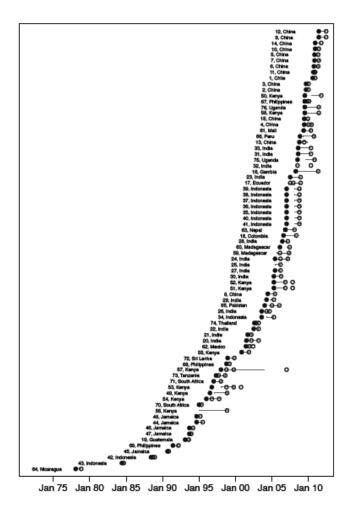
Q1: Are more studies being conducted?

Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments

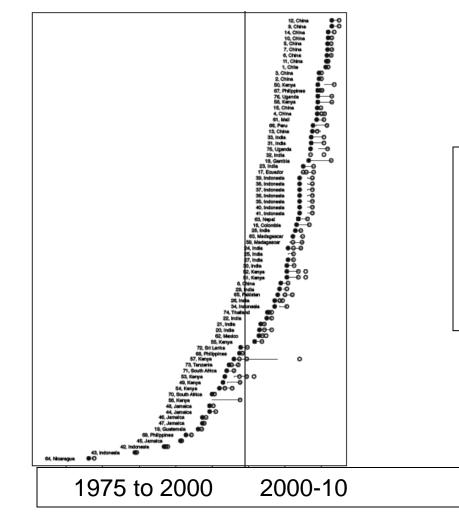
> Patrick J. McEwan^{*} Wellesley College

> > May 2013

Dates of treatment and data collection in 76 "learning" experiments



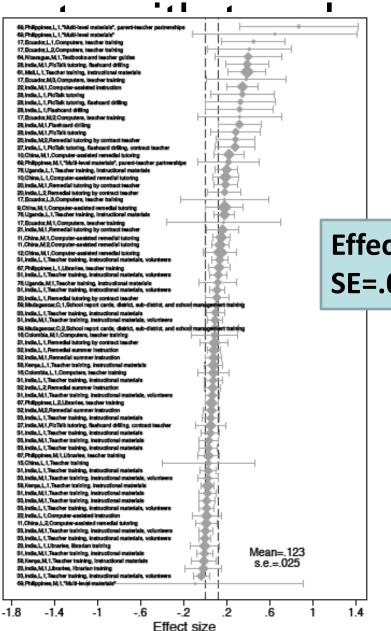
Dates of treatment and data collection in 76 experiments



Twice as many studies on learning in the ten years after 2000 ... than in the 25 years before 2000

2	nda ●0 da ●0 mya ●00 mya ●00 mya ●00
<2000 2003 2006	2009 2012

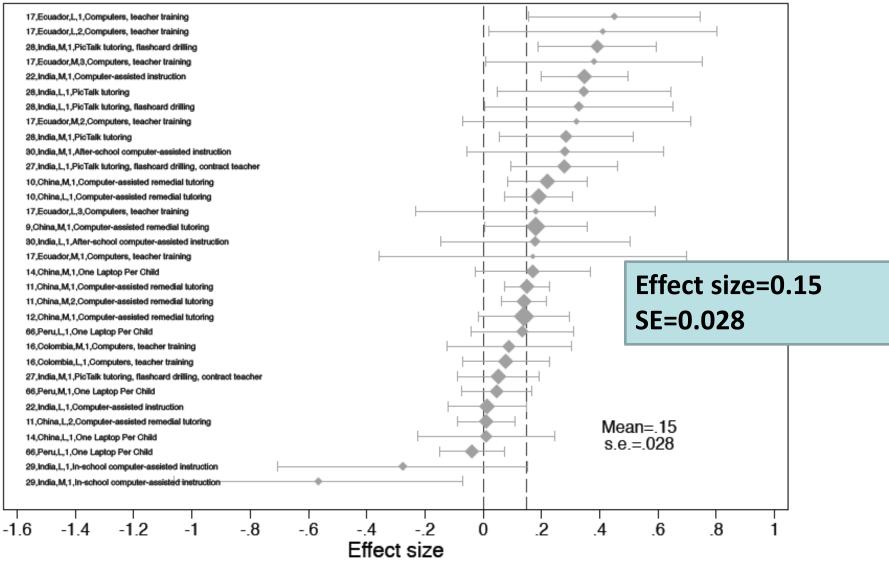
Treatm



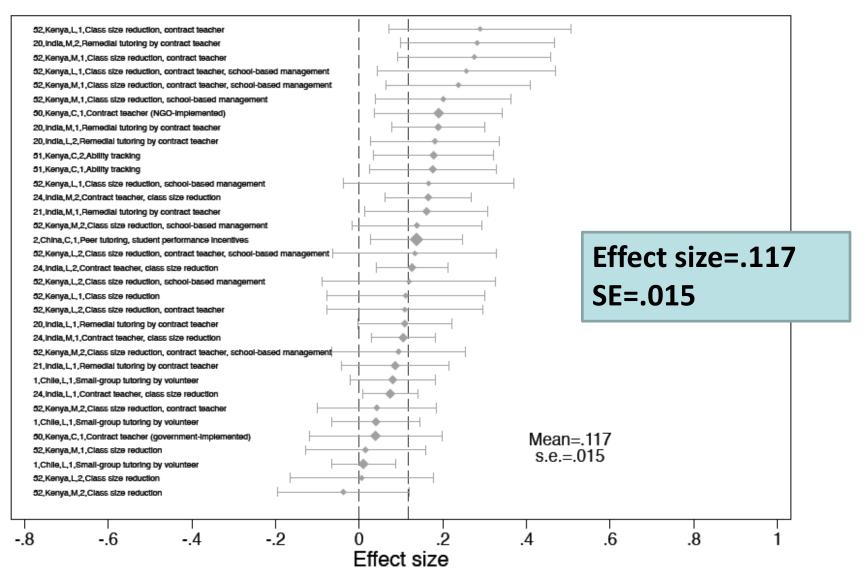
r training

Effect size=.123 SE=.025

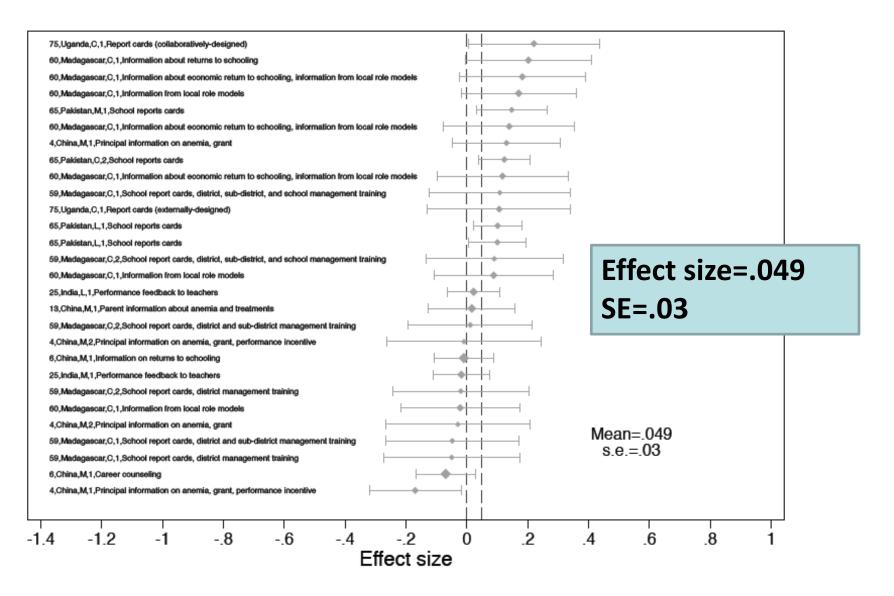
Treatments with computers or technology



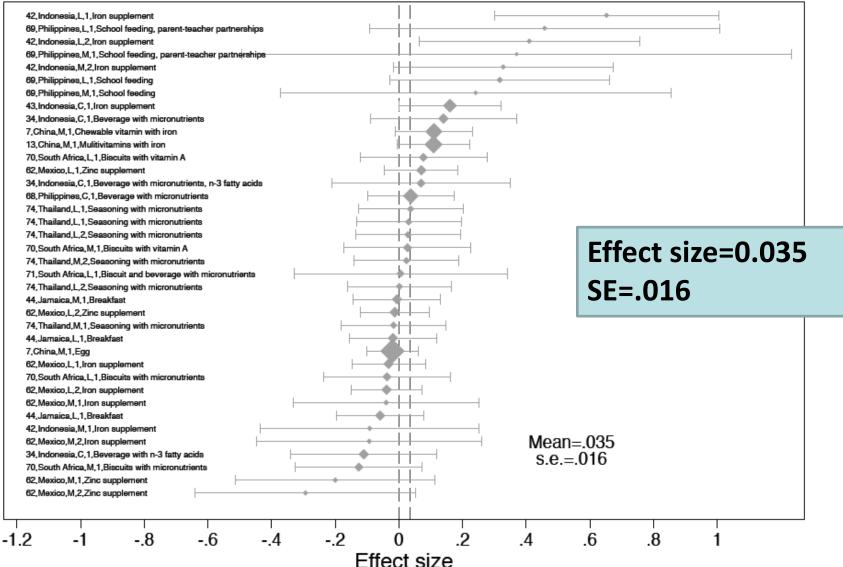
Treatments that modify class size, small-group instruction, or group composition



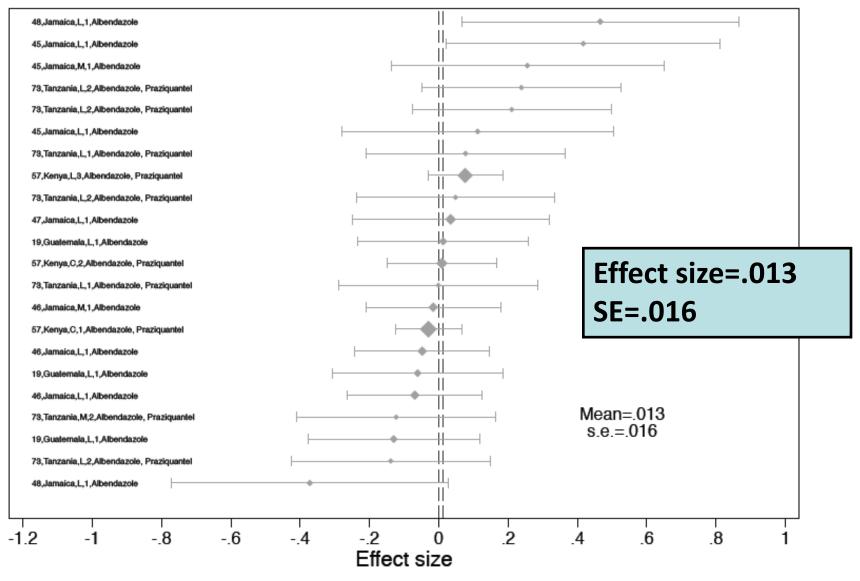
Treatments with information



Treatments with food, beverages, and/or micronutrients



Treatments with de-worming drugs



Are more studies being conducted? Summary

- A lot more studies are being conducted ...
- Accelerating rate ... [By the way: significant share supported by 3ie]
- We are beginning to learn about what works and what does not
 - Average size of effects are zero to small to medium sized
 - Large heterogeneity → by country / by type of intervention
- What works to bring kids into school / does not necessarily work to improve learning

Q2: How are the methods that are utilized changing over time?

Our first experiment in China

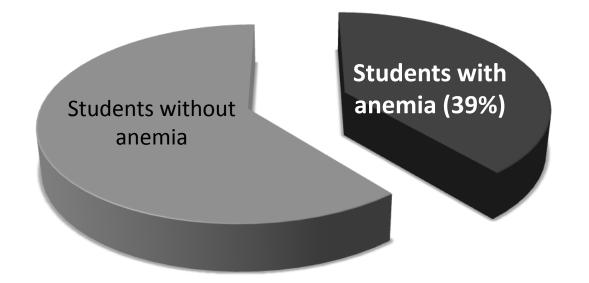
Pilot study: 2007

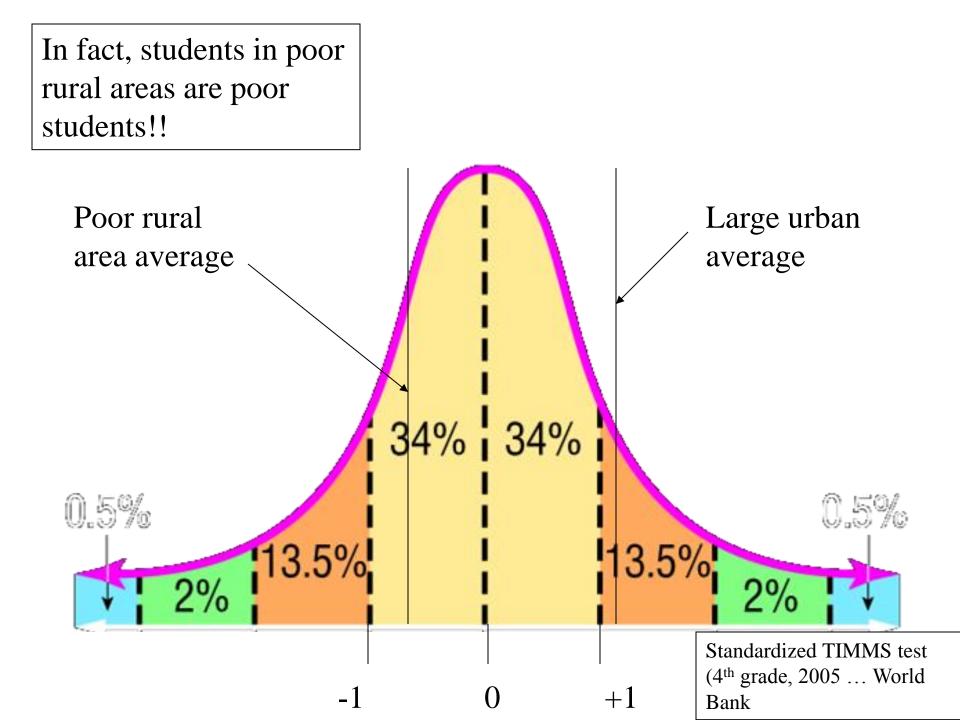
- Conducted a survey of 4158 fourth grade students in nine counties in rural Shaanxi province
- Tested students' hemoglobin levels



Initial Results

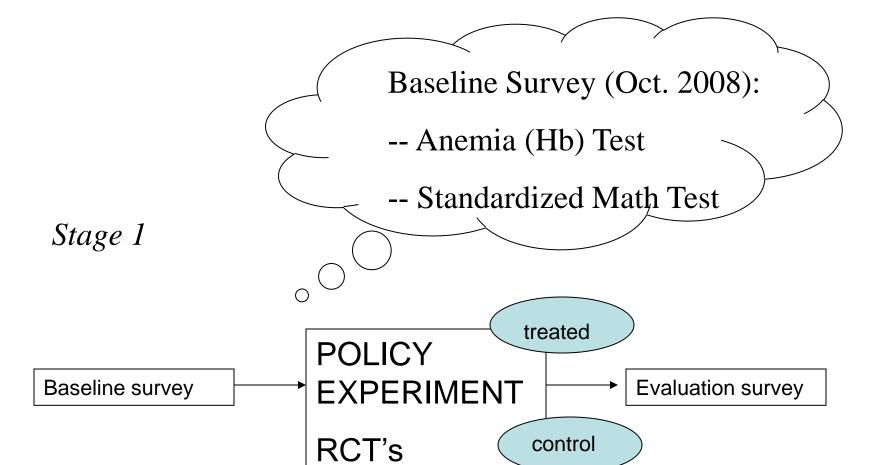
• Found that 39% of students had anemia

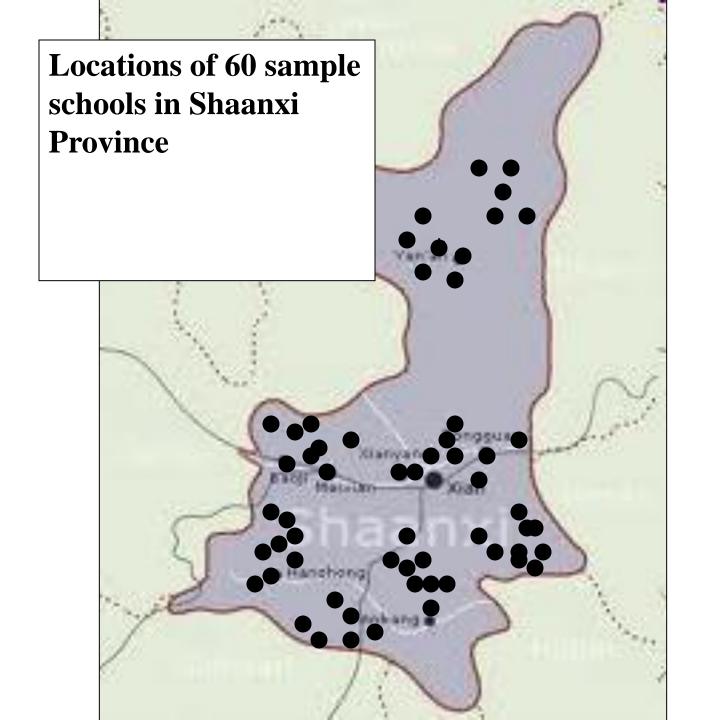




What can be done? Will iron supplements → less anemia?

Lower anemia -> Better school performance?



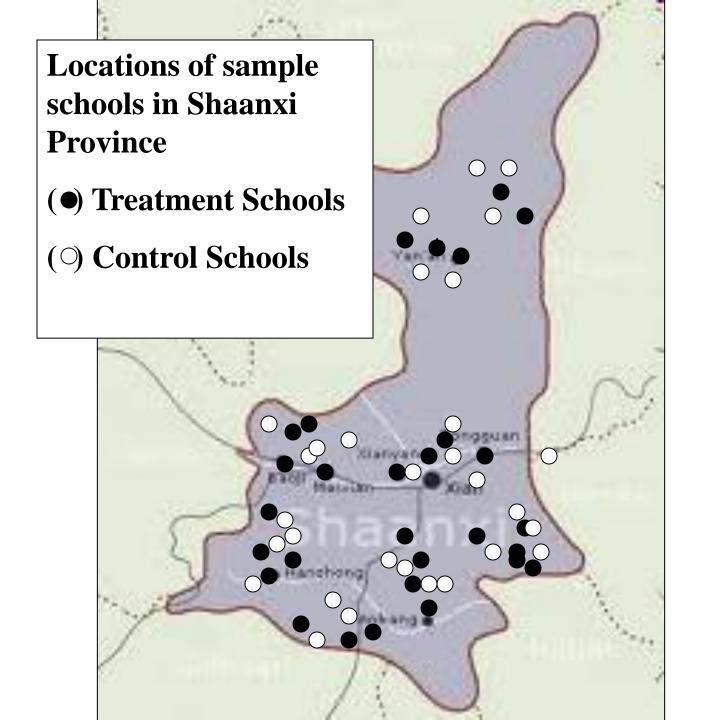


Using Hemocue 201+ technology gives Hb levels in 45 seconds (Oct. 2008)



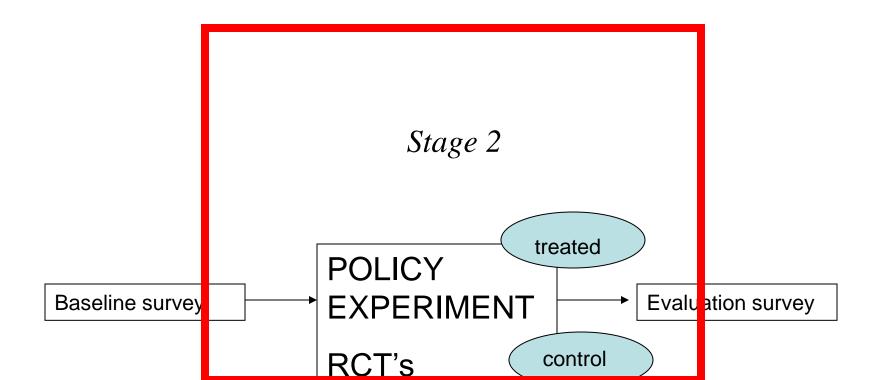
Baseline TIMMS test (October 2008)

All fourth grade students



Stage 2 → The Intervention Will iron supplements → less anemia?

Lower anemia → Better school performance?



The Intervention

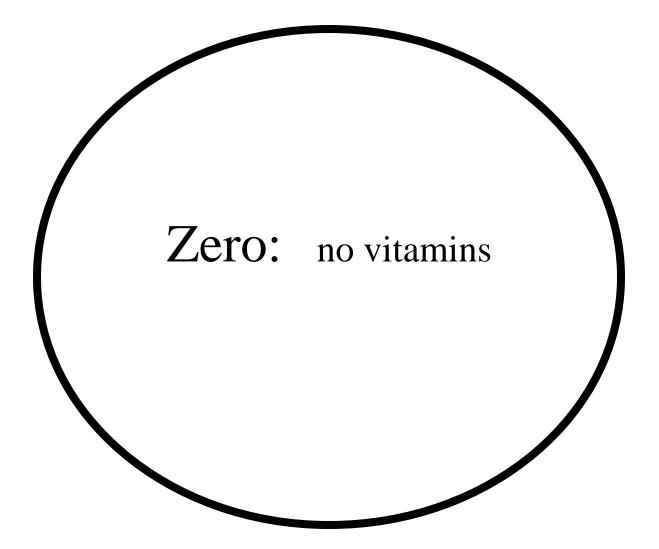
- School Type A (30 schools)
- "Centrum / Day"
- Give students one multi-vitamin with iron per day (5 mg of iron) ... from November 2008 to May 2009 (≈4 US cents/day)





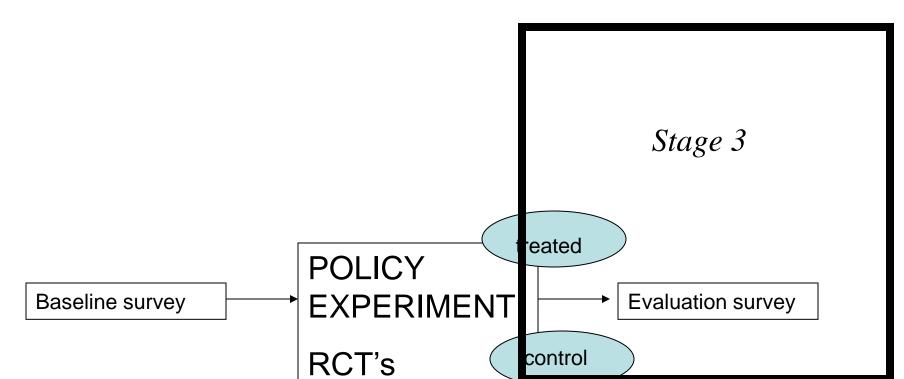


30 control schools



Evaluation Survey (stage 3) Will iron supplements → less anemia?

Lower anemia → Better school performance?



Evaluation survey (June 2009)

Re-taking the standardized academic tests ...

... after 5 months of vitamins + letter to parent.

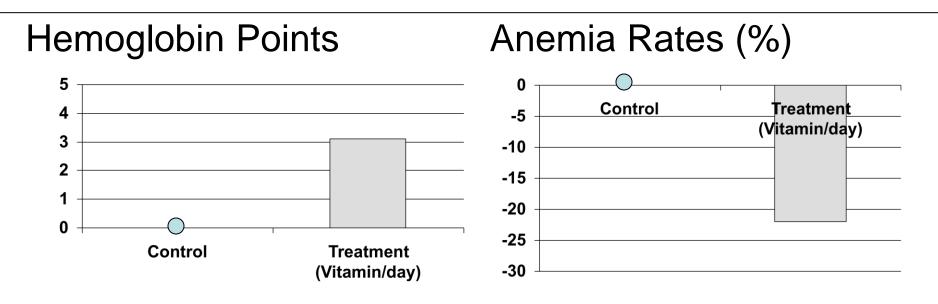


Re-taking the Hb Test (June 2009) Still anemic or not?

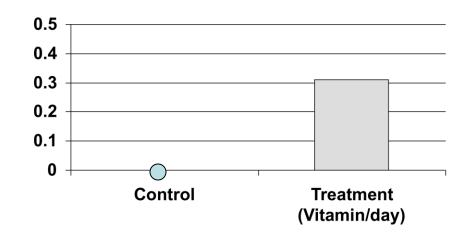


Results

Impact of vitamin on students:



Math Test Scores (std. dev.)



Great results ...

Publishable (of course: we got statistically significant results)

Policy impact (??)

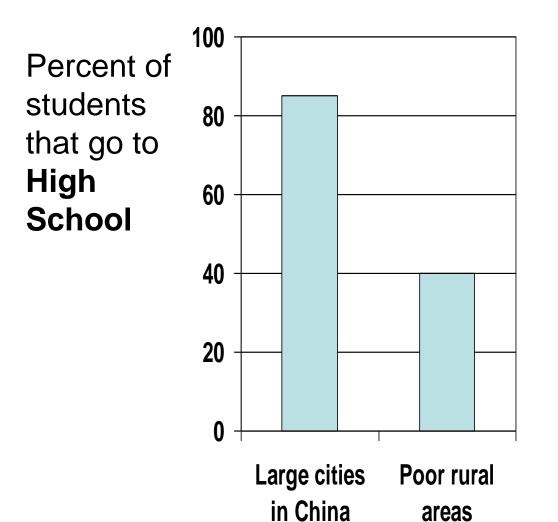
- -- who did it effect?
- -- compliance?
- -- external validity?
- -- cost effectiveness?

But, even "worse"

What happens if something like this happens?

The PROBLEM: High School Gap in China today

China in the 2010



Reasons

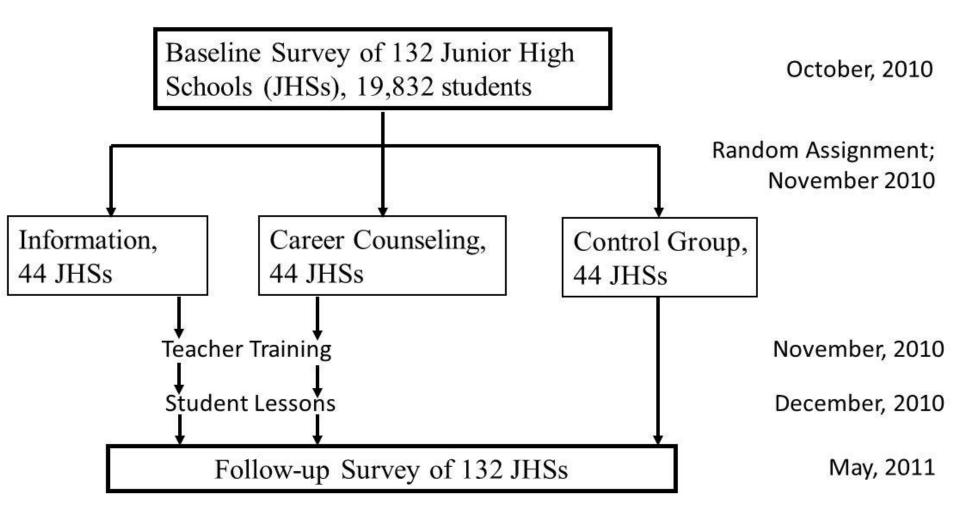
- **\$\$\$\$** ????
 - Yes → but, decentralized fiscal system → local government will not spend \$\$\$\$ on making high school free → more educated people are, faster they leave the county!!

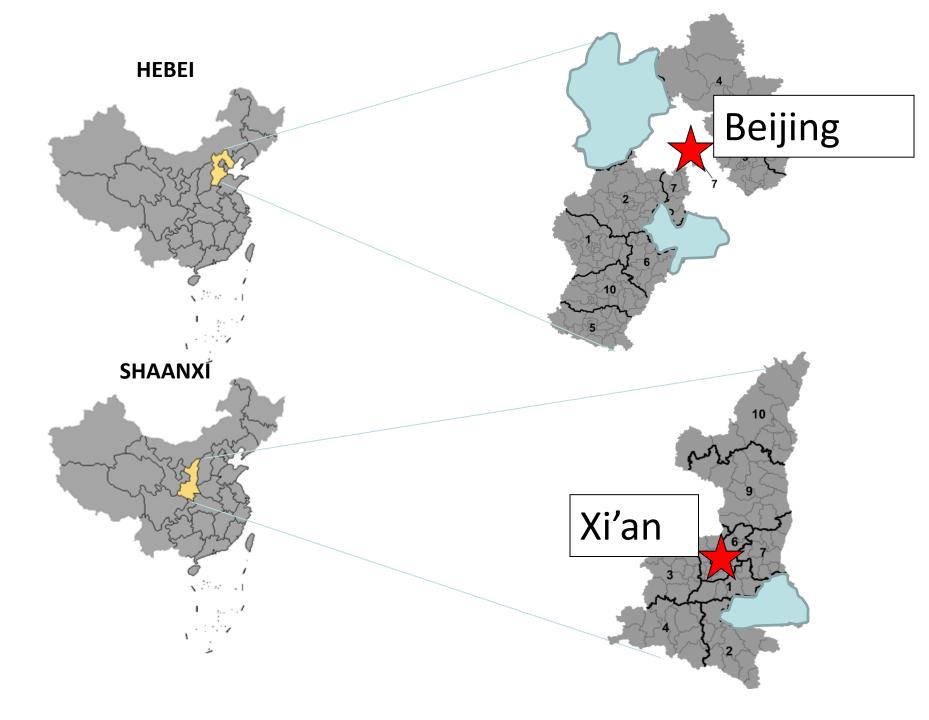
... or ...

- Information ????
 - High school conveys high benefits on families ... and will likely be even higher in the future
 - And, it is "inexpensive"

→ "Let's try this!!"

Step 1: Design





Step 2: Baseline Survey (2010.10)

- Student survey:
 - standardized math test;
 - household assets;
 - family characteristics;
 - schooling characteristics.
- Teacher and school survey





Students filling in student forms. [Right above]

Teachers filling in teacher forms. [Left below]

Step 3: Interventions (2010.12)

Information Intervention:

Professional counselor trained teachers to give a scripted 45 minute lesson on the wages and costs associated with different levels of schooling

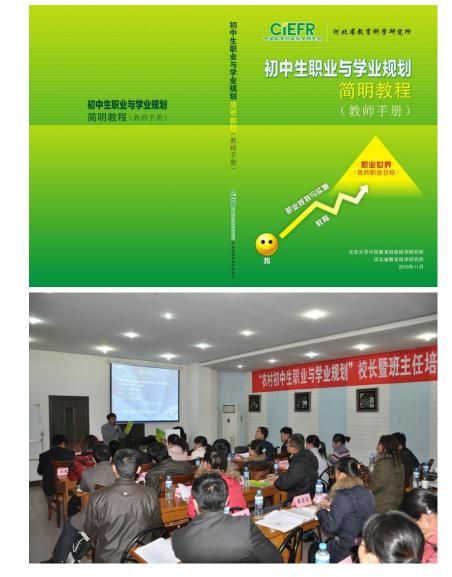
Career Counseling Intervention

Professional counselor trained teachers to give 4 scripted lessons (45 minutes each) on:

a. The world of work & your place in it

b. Careers overview & career planning

c. The information intervention



Step 4: Evaluation Survey (2011.05)

- Near identical survey forms as the baseline:
 - Standardized math test
 - Student forms
 - Teacher forms





Special attention given to identifying the dropped out students (distinguishing them from those who transferred out, repeated a grade or were temporarily absent).

Results

Results: Information and Counseling

	(1)	(2)	(3)	(4)	(5)
	dropout	math 2011	any HS	acad. HS	voc. HS
information	.01	01	01	03	01
	[.01]	[.04]	[.01]	[.02]	[.02]
counseling	.01**	07	.00	.02	.02
	[.01]	[.04]	[.01]	[.02]	[.02]
Ν	12,082	10,848	10,837	10,837	10,837
R^2	.08	.36	.14	.17	.29
Robust standard errors in parentheses, *** p<.01, ** p<.05, * p<.1 Covariate-adjusted analyses					

Impacts of:

- Information:
 - On drop out 🗲 none
 - On plans for high school \rightarrow
 - One learning -

none none

- Counseling: – On drop out 🗲 more drop out!! - On plans for high school \rightarrow none none
 - One learning -

Which journal?

JNR [journal of non-results]

Which policy maker?

[who wants to push something that failed?]

But, that was before ...

- In response to critique of Impact Evaluation's focus on: "impact/no impact" only ...
- ➔ Theory of Causal Chain Analysis

CCA helps address one a major criticism of the current wave of IE studiesThey only tell us: what works and not much else!

- If research teams can address this criticism, help answer a number of challenges:
 - Why would anyone want to be told that their project does not work
 - World Bank employee?
 - Government official?
 - NGO? Teach for China
 - If you only know it does not work, what is the implication? Eliminate the program ... or fix it? But, how? How to make it work better?

International Initiative for Impact Evaluation (3ie) is an international organization trying to put the "how" in rigorous IE with "theory-based evaluation" or "causal chain analysis"

Example: a nutrition project in Bangladesh

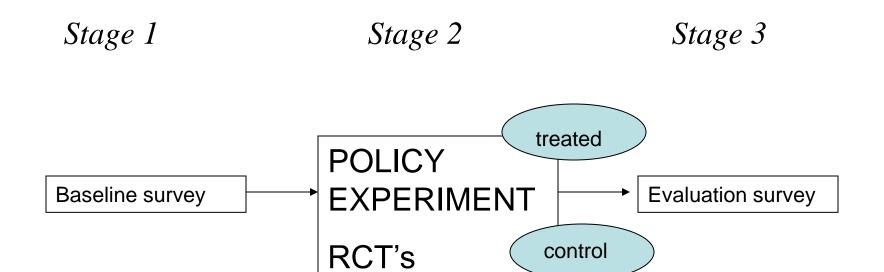
 Source: Howard White and Edoardo Masset (2007) 'The Bangladesh Integrated Nutrition Program: findings from an impact evaluation' *Journal of International Development* 19: 627-652

- Bangladesh Integrated Nutrition Project (BINP) ... a World Bank Project
- Problem: lots of malnutrition ... difficult to solve in traditional institutional structures

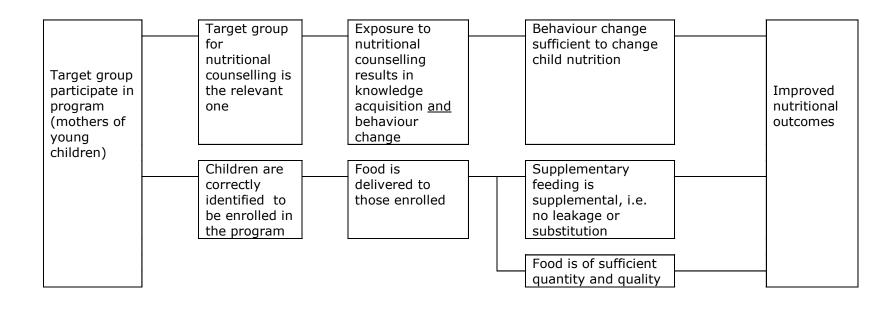
 \rightarrow

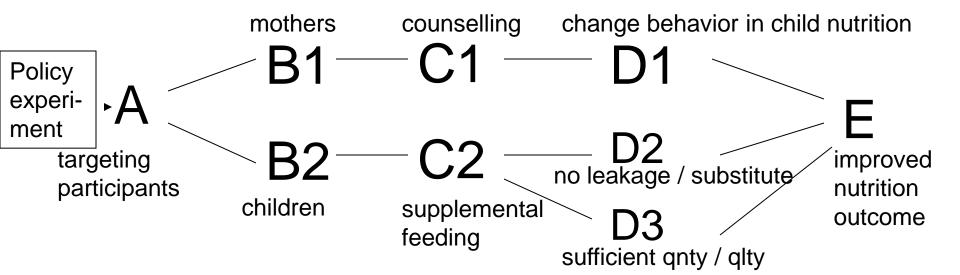
- Growth monitoring, nutritional counselling and supplementary feeding (based on a program in Tamil Nadu, which was successful)
- According to the design of the project, implemented by NGOs at field level, used Community Nutrition Practitioners (CNPs)

Instead of: Simple Search for Impact Will MCH program → improve nutrition?



Program design (theory of change)



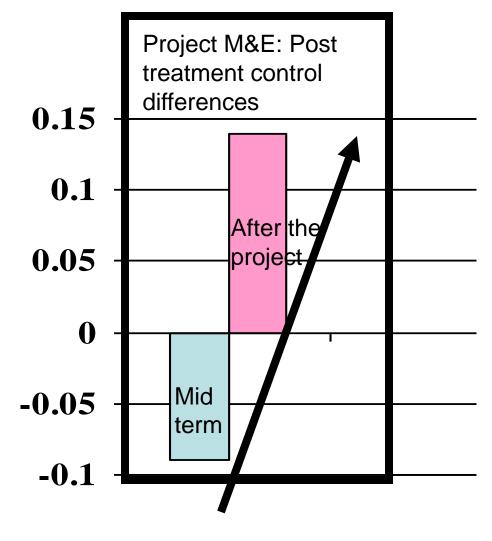


The evaluation story

- Looked like it was working all bits in place and outcome monitoring data showed fall in severe malnutrition
- Bank agreed to scale up (this is an expensive program ... funded at expense of other projects)
- But Save the Children UK critical, though Bank's M&E team was positive
- Bank's evaluation department (IEG) did a more rigorous evaluation found little or no impact
- Theory-based approach explains why

Measuring outcomes and impacts (M&E)

Height for Age Scores single differences (between treatment and controls)

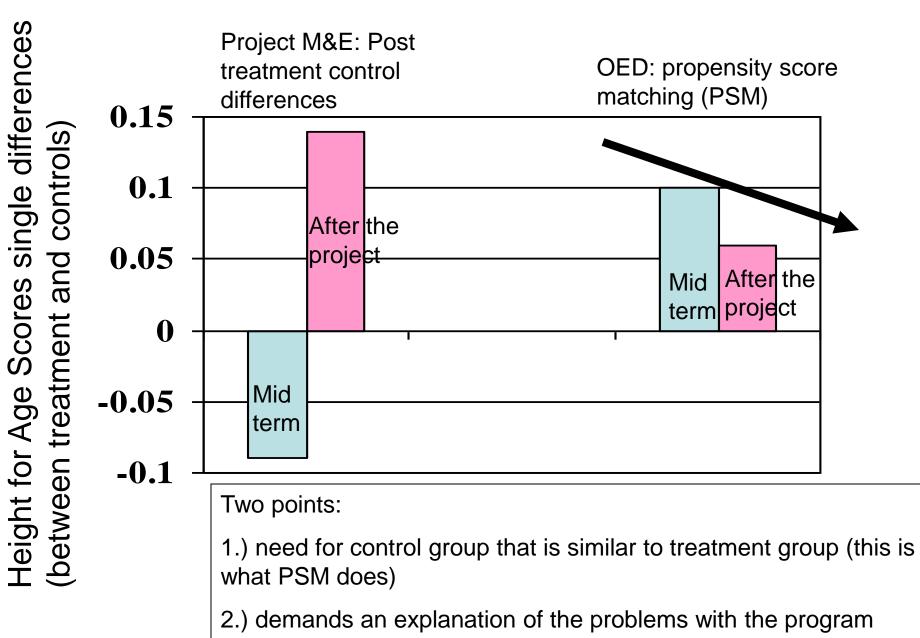


The evaluation story

- Looked like it was working all bits in place and outcome monitoring data showed fall in severe malnutrition
- Bank agreed to scale up (this is an expensive program ... funded at expense of other projects)
- But Save the Children UK critical, though Bank's M&E team was positive
- Bank's evaluation department (IEG) did a more rigorous evaluation – found little or no impact

Theory-based approach explains why

Measuring outcomes and impacts



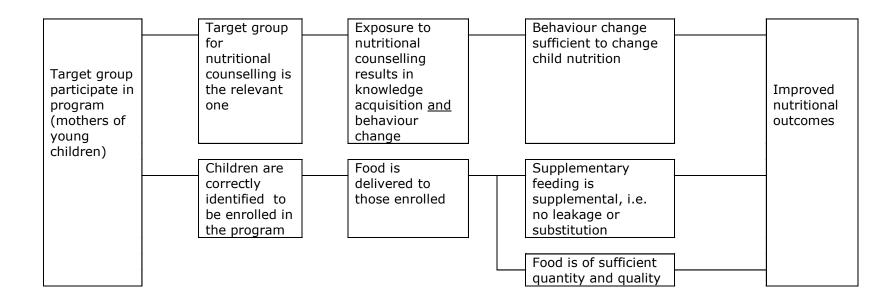
The evaluation story

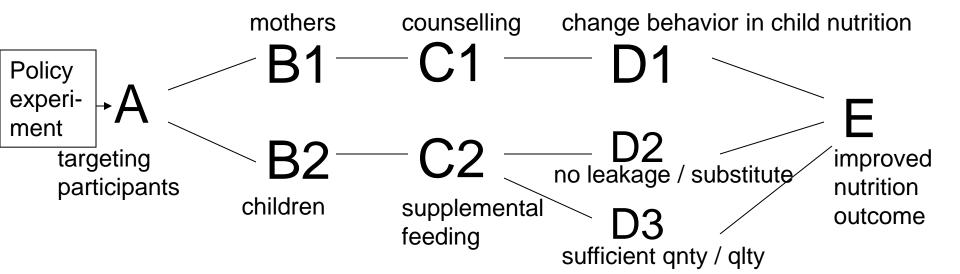
- Looked like it was working all bits in place and outcome monitoring data showed fall in severe malnutrition
- Bank agreed to scale up (this is an expensive program ... funded at expense of other projects)
- But Save the Children UK critical, though Bank's M&E team was positive
- Bank's evaluation department (IEG) did a more rigorous evaluation found little or no impact
- Theory-based approach explains why

Implementing theory-based analysis

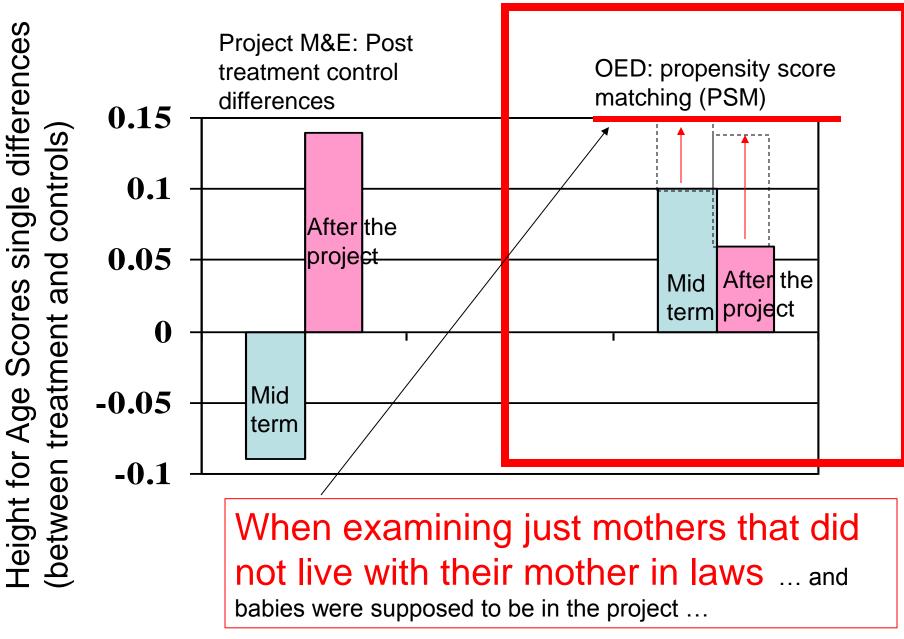
Assumption	Findings
3. Provide nutritional counseling to care givers	Mothers are <u>not</u> decision makers, especially if they live with their mother-in-law
2. Women know about sessions and attend	d 90% participation, lower in more conservative areas
1. Malnourished and growth faltering children correctly identified	g No – community nutrition practitioners (CNPs) cannot interpret growth charts
4. Women acquire knowledge	Those attending training do so
5. And knowledge is turned into practice	No there is a substantial knowledge- practice gap
6. Supplementary feeding is additional food for intended beneficiary	No, considerable evidence of substitution and leakage
Adopted changes are sufficient to improve intended outcomes	Only sometimes → for mother/caregivers

Program design (theory of change)





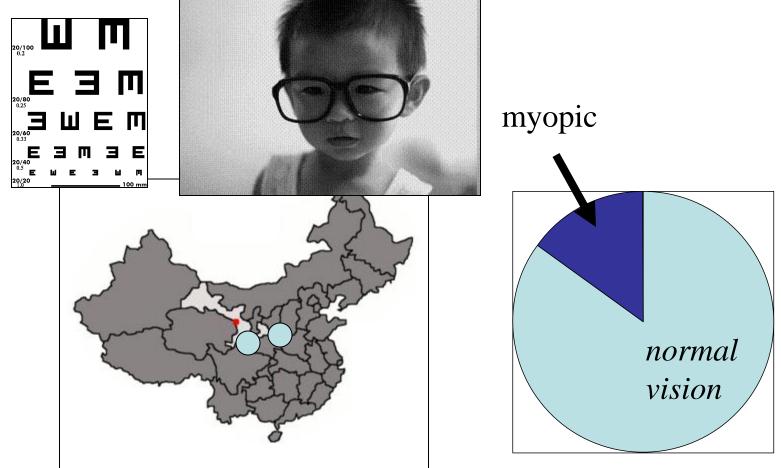
Impacts when mother participated



How to implement "Theory of Causal Chain"?

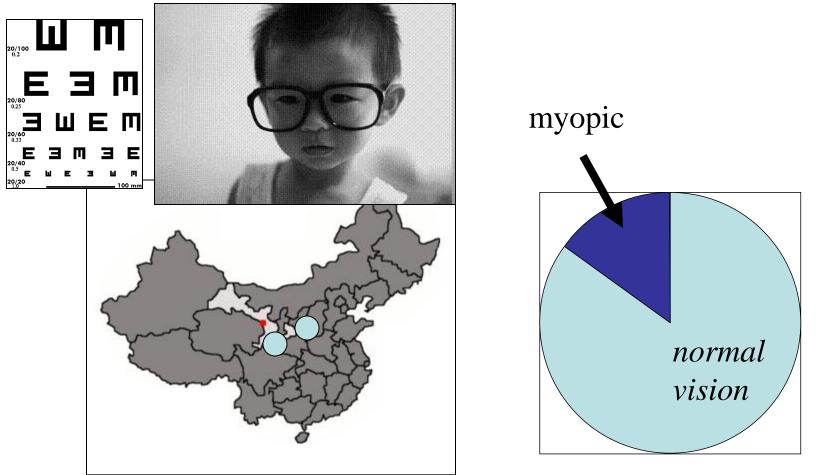
- It needs to be developed BEFORE / AS PART OF → the Project Design ...
- It is the plan around which all activities are based:
 - [Preanalysis Plans]
 - Sampling / Identification of Outcome
 - Baseline Survey
 - Design of the Intervention
 - Endline Survey
 - Analysis

Testing 19,500 children in Gansu and Shaanxi Provinces



 \rightarrow 5000 (\approx 25%) were myopic (or nearsighted).

Testing 19,500 children in Gansu and Shaanxi Provinces



→ 5000 (\approx 25%) were myopic (or nearsighted).

Only 650 had eyeglasses (≈ 3%)

Main Question of Project

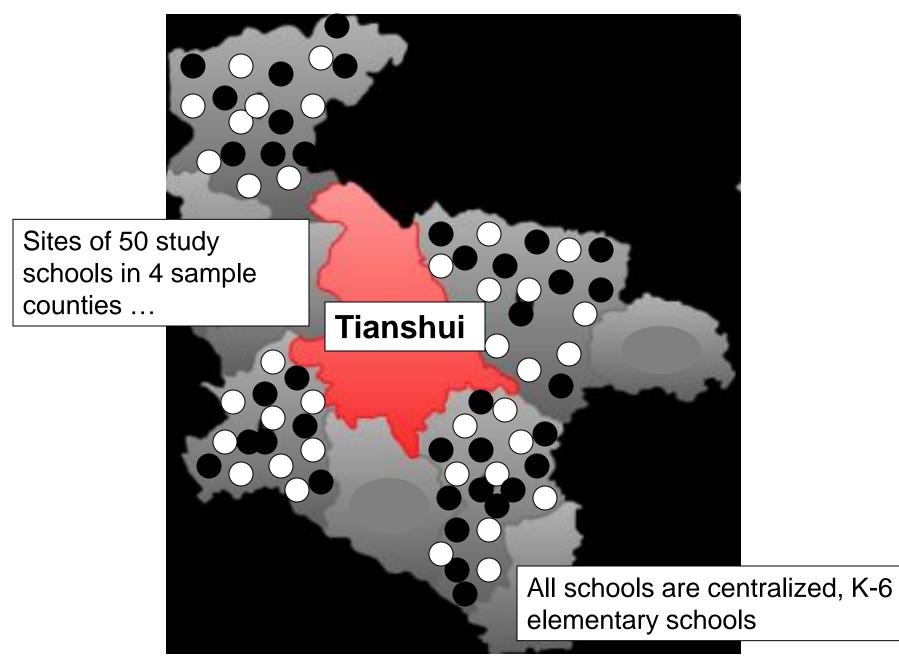
 Can we improve educational performance by providing eyeglasses to students?

Two approaches

- Free glasses / exam in schools / dispense in school
- Vouchers / exam in schools / make family + student undertake an ordeal (up to 4 hour bus ride) to pick up glasses

interve provide uptabe mediating glasses 9 gas ke ala Sou Tracher sunk 1007. Sunda Cot In-clar cost Seat Kills NO Deler. In-house rules teased. It incluse actuates feasing ula 1/1st. Actual G Kido Dinnet bor tool Col supp - distance - Knowledge time no ; mild; nul; severe parent - Vision problems - experience / glasses trau Por -00 unconfiel tutson an souly che - te color attention - solf estern - ad

Affects sampling ...





No.	
2. What supers a most of year land income of	
 Beller for an end for an load of proving well-been or even that the set back of the set of the set. 	In such a second
 A. of probability second pairs for instrument in processing and the probability of the probabil	P
 When is the maximum Weil (WHIP) per scattering to advance on any little 	a honait seculture, 7
 West and go the state processing assess 47 here some and somethy Parents, Adda 	of endingene south (m.
1. Now be was considered with tail the ingle provide counds	agone and data would be
Prover bet when a produced we do that you might prove the second se	f Aire
Prover between a produced within total true implementation and the second	(Are P
Prove for venue - produced view todal transmitter research Transmitter (Transmitter Provide Transmitter Provide Trans	A set of barres and be

1. Print cards your restores to add emunicipality.

Maryaty ter	Arter		Caratra -	Neger	Loung's
Carryonity, I do NFT have compt shout the existence of emitylay we is consistent to machine applies on no hard	3.1	11	14	14	18
The planning period for well-dipute will conflict with the planning period for sim- when usings	1	1	41	200	3.9
The herr rules partial for weighterne, will collisi with the harvesting partial for ne- effor usage	4.	$\langle t \rangle$	0		L.A.
Markets for excitingness are SOT fit-shaped acception afrest as increase	1	1.	1	1.6	
Frederican sale for code by too is former. Base the other straps I produce	1.	1.1	10	1	
I would consider signifig a long-term extended provided Presend with a company in grow with layout for charge	1.1	4	- 10	2.6	1.1
I would and industrial permission in home have in given and knowed could known	1 (L	-36	. 0.		
Her aphen ander is produce under deury: Best-Instance (Her weichtigtung in bereine Beite zweige independent		1	- 4		
I must fin is provid man balance for	- T -	1.1	10		
I would need providence permitted to produce which gives	1	t	1.1	1.4	

Affects survey design ...

... and analysis ...

- ... with solid Causal Chain Analysis, can do the following:
- Does project work or not?
- Who does it work for?
- If it does not work, why does it not work?
- If it does work, what could make it work better?
- Answer a lot more questions that policy makers are interested in ...

Q3: Are study findings making their way into policy decisions in developing countries and at organizations like the multilateral development banks? In part in response to the availability of evidence-based research, China has created a system for connecting research (evaluators) with the top leadership



2009 第 14 期(总第 84 期) 中国科学院农业政策研究中心 2009 年 12 月 1 日

减少贫困地区儿童贫血现象,促进儿童全面发展

儿童营养与健康是关系到国家未来发展的重要因素之一。联合国儿 童基金会的研究表明,发展中国家有一半左右的儿童患有贫血,其中铁 元素摄入不足和吸收不充分是最主要原因。儿童贫血不仅影响身体健康, 而且会导致注意力集中时间短,学习能力差,长期来看会影响德智体的 全面发展。为了了解我国贫困地区儿童贫血情况,中科院农业政策研究 中心农村教育行动计划课题组于 2008 年 11 月在陕西省 8 个贫困县 66 个 小学,对 4000 多名农村学龄儿童(四年级学生)的贫血和学业表现状况 进行基线调查,并对他们进行了为期半年的营养补充和家长营养知识宣 传试点干预;在此基础上,从 2009 年 6 月开始对试点干预结果做了评估 和分析,期望通过上述干预研究探索出解决贫困农村儿童贫血的可能办 法。以下是我们调研的主要结果。

一、貧困农村儿童貧血状况不容忽视

 貧困地区儿童貧血比例超过 1/3. 按照我国同年齡段貧血诊断的 标准(血红蛋白水平低于 120 克/升),基线调查的检测结果表明陕西貧 困地区四年级的学生有 38%存在貧血。虽然这一结果比其它发展中国家 的平均数要好,但据此结果估算,全国貧困地区还有上千万的儿童存在 不同程度的貧血。

 寄宿学生貧血情况更为严重,基线调查结果表明,从貧血比例来 看,寄宿生(41%)显著高于非寄宿生(36%);中午在学校吃饭的学生(41%)

Official policy brief (think of President Obama's desk and the desks of his cabinet members)

中国科学院

中国科学院政务信息采用通知

中科院农业政策研究中心:

感谢贵单位对我们工作的大力支持,希望继续关注和支持我院的 政务信息工作,及时将本单位的科研成果、发展动态以及针对我院和 国家的重要建议等重要信息报送办公厅,为国家发展和我院的知识创 新工作做出更大贡献。

联系人:万,万圆

联系电话: 010-68597568

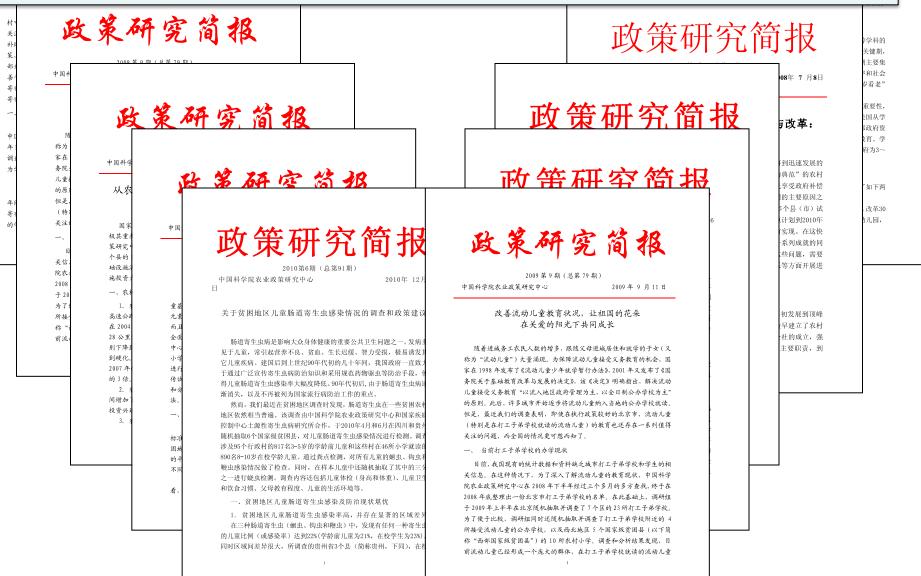
E-mail : shishuo@cashq.ac.cn

Formal notification that there has been a "policy directive" directing MoEdu and MoHealth to move our information into the

nolicy discussion

政策研究简报	中国科学院
2009 第 14 期 (总第 84 期) 中国科学院农业政策研究中心 2009 年 12 月 1 日	中国科学院政务信息采用通知
减少贫困地区儿童贫血现象,促进儿童全面发展	中科院农业政策研究中心: 你单位 <u> 乾林秀、黄季焜、刻承芳、罗仁福</u> 等同志于 <u>2009</u> <u>年12月8日</u> 向办公厅信息宣传处报送的信息 <u>中科院专家关于减少</u>
儿童营养与健康是关系到国家未来发展的重要因素之一。联合国儿 童基金会的研究表明,发展中国家有一半左右的儿童患有贫血,其中铁 元素摄入不足和吸收不充分是最主要原因。儿童贫血不仅影响身体健康, 而且会导致注意力集中时间短,学习能力差,长期来看会影响德智体的 全面发展。为了了解我国贫困地区儿童贫血情况,中科院农业政策研究 中心农村教育行动计划课题组于 2008 年 11 月在陕西省 8 个贫困县 66 个 小学,对 4000 多名农村学龄儿童(四年级学生)的贫血和学业表现状况 进行基线调查,并对他们进行了为期半年的营养补充和家长营养知识查 传试点干预;在此基础上,从 2009 年 6 月开始对试点干预结累做了评估 和分析,期望通过上述干预研究探索出解决贫困农村儿童贫血的可能办	《中国科学院考报信息》2009 年第 217 期 采用, 并上银 中办 和 函办, 已被 中办 刊物 《考報》采用, 函办 刊物 《送阅》 米用, 开得到 温客 信息理、考 息 後 副 是 理, 副 办 刊物 《送阅》 米用, 开得到 温客 信息理、考 息 後 副 是 理, 副 办 刊物 《送阅》 网络谢贵单位对我们工作的大力支持,希望继续关注和支持我院的 政务信息工作, 及时将本单位的科研成果、发展动态以及针对我院和 国家的重要建议等重要信息报送办公厅,为国家发展和我院的知识创 新工作做出更大贡献。
	Li Keqiang Liu Yandong
 貧困地区儿童貧血比例超过 1/3. 按照我国同年齡段貧血诊断的 标准(血紅蛋白水平低于 120 克/升),基线调查的检测结果表明陕西貧 困地区四年级的学生有 38%存在貧血。虽然这一结果比其它发展中国家 的平均数要好,但据此结果估算,全国貧困地区还有上千万的儿童存在 不同程度的貧血。 寄宿学生貧血情况更为严重、基线调查结果表明,从貧血比例来 看,寄宿生(41%)显著高于非寄宿生(36%);中午在学校吃饭的学生(41%) 	E—mail : shishuo@cashq.ac.cn
Official policy brief (think of President Obama's desk and the desks of his cabinet members)	Formal notification that there has been a "policy directive" directing MoEdu and MoHealth to move our information into the
	policy discussion

REAP has submitted 17 policy briefs to State Council



Policy Action

These policy directives directly respond to our brief and are signed by:



But, it does not happen without a concerted effort (pre-thought out plan) and a lot of effort

- We start with P.A.C.'s [not political action committees, like here in DC]: Policy Action Committees
 - 2 to 5 local officials that will be our "advisors":
 - Get input during experimentation design ...
 - More get to understand the project
 - E.g., in the case of RCT's ... truly randomly assigned
 - If successful, PAC-members are given "ownership" and have an incentive to try to upscale ...
 - They help us push policy ... and often sign Policy Briefs ...
 - Policy Briefs can "give political cover."

Policy suggestions (with the support of a lot of other/complimentary work—of course) → associated with:

- 20 billion dollar "nutritious lunch" program
 26 million children x 200 lunches per year x
 3.5 yuan per lunch x 10 years
- New financial aid fund for poor kids at high school level
- Deworming campaign in 62 poor counties

Ironic thing: can't run a DiD regression and show causal impact of our impact evaluation ... but, the correlations are strong ...

Thank You!



http://reap.stanford.edu