

## Gold Standard? The Use of Randomized Controlled Trials for International Educational Policy

NATHAN M. CASTILLO AND DANIEL A. WAGNER

*Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty* by Abhijit V. Banerjee and Esther Duflo. New York: PublicAffairs, 2011. 303 pp. \$26.99.

*Making Schools Work: New Evidence on Accountability Reforms* by Barbara Bruns, Deon Filmer, and Harry A. Patrinos. Human Development Perspectives. Washington, DC: World Bank, 2011. 268 pp. \$35.00.

*More Than Good Intentions: How a New Economics Is Helping to Solve Global Poverty* by Dean Karlan and Jacob Appel. New York: Dutton, 2011. 308 pp. \$26.95.

Edward Miguel and Michael Kremer pioneered a new kind of development research in their 2004 study of a school deworming program in Kenya. Their experimental design incorporated the random assignment of primary school students to either a treatment or a control group for receiving medicine to eliminate intestinal parasites. Findings revealed significant benefits to the treatment group in not only improved health but also lowered school absences (Miguel and Kremer 2004). One policy consequence was an increased awareness for more evidence-based decision making under the banner of accountability reform in international development.<sup>1</sup> The driving focus for such reform is rigorous scientific investigation—what some call the “gold standard” of methodology—that uses randomized controlled trials (RCTs) to establish a credible link between an intervention and a set of outcomes.

Increasingly, major agencies and specialized evaluation committees are pressuring the field of education development to include such methods as a way to buttress scientific and policy arguments. This review takes a closer look at three recent publications that present complementary versions of this evidence-based (EB) perspective to identify strengths and limitations for making policy decisions in low-income countries (LICs). While the increased use of empirical and credible results is always welcome toward the broad goal

<sup>1</sup> Over the years, a number of organizations, such as the US Department of Education’s What Works Clearinghouse, the Campbell Collaboration, International Initiative for Impact Evaluation, and the Coalition for Evidence-Based Policy, have pioneered efforts in identifying, assessing, and promoting evidence-based research for policy making within the domain of education.

*Comparative Education Review*, vol. 58, no. 1.

© 2013 by the Comparative and International Education Society. All rights reserved.  
0010-4086/2014/5801-0007\$10.00

of improving education quality, the degree to which EB findings are generalizable for useful policy purposes in diverse societies is much less clear.

#### “Controlled” Evidence Meets “Real World” Impact

Within the domain of science, credible evidence is predicated on the notion that other factors beyond the parameters of the study are not significantly influencing the observed outcomes. In the field of medicine, researchers produce this kind of controlled experimental environment by randomly assigning one group to receive a particular treatment of interest while another group does not. All other conditions being equal, researchers can confidently make powerful causal claims based on their observations. While producing a truly controlled experiment within the context of social sciences is much more complicated, the practice of randomly assigning groups to either receive an intervention (treatment group) or not (control group)—randomized controlled trials—has become increasingly popular among education planners and policy makers.

The first volume reviewed here is *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*, by Abhijit V. Banerjee and Esther Duflo. One major claim in *Poor Economics* is that “small changes can have big effects” (272). Further, the philosophy presented throughout the book—from improving politics to “reengineering education” (97)—is to “attend to the details, understand how people decide, and be willing to experiment” (253). In 10 chapters, the authors rely on an 18-country data set to provide empirical evidence for what works and what falls short of success in international aid. The research is presented through a series of vignettes that reintroduce (following in the footsteps of decades of anthropologists) the notion of the rationality of poor people’s behaviors, even if sometimes counterintuitive to economists.<sup>2</sup>

Banerjee and Duflo refer to the supply versus demand debate to illustrate the nuances of predicting human agency within the domain of poverty alleviation. Typically, the supply-wallahs (e.g., Jeffrey Sachs) emphasize greater access while the demand-wallahs (e.g., William Easterly) advocate increasing the demand for quality education (77).<sup>3</sup> Banerjee and Duflo’s view is that parents often do not have sufficient information as to what the real returns on education are and therefore do not know how to respond to external (usually governmental) interventions. Banerjee and Duflo argue that supply versus demand approaches are more complementary than oppositional. Accordingly, they contend that practices should combine strategies to work within particular contexts where local belief systems can be understood.

<sup>2</sup> See, e.g., Aikman (1999); Clammer (2012).

<sup>3</sup> The authors make reference to India’s Green Revolution and the recent proliferation of offshore call centers as explicit examples of this argument.

In some development contexts, governments have adopted a policy tool that has become known as a conditional cash transfer (CCT) to encourage parents to make the investment for their children by subsidizing family income. The authors present a well-known and highly publicized educational incentive scheme implemented in Mexico (initially called *Progressa*; later called *Oportunidades*). This project was designed to encourage investments in human capital by linking family welfare payments to school matriculation and preventative healthcare. A government-sponsored RCT study showed that the program had a substantial impact on enrollment and dropout prevention, particularly at the secondary level (79).<sup>4</sup> However, the growing number of global CCT studies suggests that outcomes are susceptible to context variations and present concerns about the generalizability of findings.<sup>5</sup>

Banerjee and Duflo present another example from an RCT study in Kenya. Twenty-five out of 100 schools were randomly selected to receive textbooks, which have been often regarded as positive inputs for educational achievement. However, the evaluation revealed inconclusive findings between treatment and control schools except for treatment students who excelled at the outset.<sup>6</sup> Banerjee and Duflo mention that while English is the official language of instruction (as well as the language in which the texts were printed), most Kenyan children in this study spoke it as a third language. To the outside observer, the finding of no impact of textbook provision for average students is unsurprising since the textbook intervention was in the wrong language for many children. Clearly, context puts a strong limitation on the generalizability of even highly skilled evaluations. As this study demonstrates, technical expertise cannot overcome shortcomings in cultural understanding.

The second book, by World Bank economists Barbara Bruns, Deon Filmer, and Harry A. Patrinos, is entitled *Making Schools Work: New Evidence on Accountability Reforms*. Using evidence from 22 rigorous impact evaluations in 11 LICs, the authors discuss three critical education reform domains: transparency campaigns, decentralized administration through school-based management, and teacher incentives. Overall, the authors focus on the challenge in service delivery termed the “principal-agent problem” (10)—the complex relationship between ministries of education (principals), their

<sup>4</sup> A replication of the CCT in Malawi (Baird et al. 2009) and more recently in Morocco (Benhassine et al. 2010) investigated the impact of the conditionality of the transfer on matriculation rates. The differences between the conditional and unconditional treatment groups were not statistically significant, suggesting that participating parents merely needed a financial subsidy in order to send their children to school as opposed to a strict mandate (80).

<sup>5</sup> In a recent assesment of 16 CCTs in Latin America, Lomelí (2008) found limited effects on long-term education outcomes and poverty alleviation. Alternatively, in a review of evidence from 13 programs, Ranganathan and Largarde (2012) found CCTs to be effective in promoting preventative healthcare and modification of risky sexual behaviors. For more on design implications of CCTs for poverty alleviation, see Fiszbein and Schady (2009).

<sup>6</sup> Glewwe et al. (2000).

agents (or service providers/teachers), and the clients (beneficiaries). The authors view the poor delivery of quality education as due in part to the labor-intensive dimension of teachers' work with diverse student aptitudes and learning styles. Bruns and colleagues present an accountability framework where the improvement of teacher quality on student learning is positively influenced by "giving parents and students a direct voice in their local schools" (12).<sup>7</sup>

School-based management (SBM) interventions represent a good example of their accountability approach—where decision-making authority of school operations is transferred to local agents. As described by Bruns and colleagues, El Salvador was the first country in the developing world to adopt SBM through the implementation of the *Educo* program in 1991 (17).<sup>8</sup> Within the *Educo* model, community-elected oversight committees received funds from the Ministry of Education to carry out school-level education policies. This effort to provide information in a more timely way for development decision making is of growing importance for international agencies.<sup>9</sup>

Another way of strengthening accountability is to focus on teachers. Bruns and colleagues highlight two types of reforms, contract tenure and pay for performance. Results from their multicountry data set show that the hiring of short-term contract teachers can have significant benefits.<sup>10</sup> In pay-for-performance schemes, Bruns et al. describe how different types of programs can influence the behavior of teachers and the sustainability of the intended change in pedagogy. They cite a study in Andhra Pradesh (India) that offers persuasive evidence of incentive-design characteristics on teachers by comparing the impact of four distinct treatment conditions (159).<sup>11</sup> According to Bruns and colleagues, incentive-based policies for teacher accountability can strengthen education quality at a fraction of the cost of "wholesale reform" (196). Institutional accountability will no doubt be an area of increasing interest over the coming years.

The third volume, *More Than Good Intentions*, by Dean Karlan and Jacob Appel, focuses on behavioral economics in international development settings. As with the previous two volumes, *Good Intentions* uses EB evaluations

<sup>7</sup> Recent research from Hanushek and Rivkin (2010) and Farr (2010) demonstrate that assignment to weak vs. great strong teachers can result in as much as a full grade-level disparity in curriculum mastery among students (cited in Bruns et al.).

<sup>8</sup> *Educación con Participación de la Comunidad* (Education with community participation).

<sup>9</sup> See Wagner (2011) on improving the speed of collecting and analyzing data from learning assessments.

<sup>10</sup> Particularly in India and Kenya, the authors discuss robust findings at as little as 10 percent of the cost of civil service teachers (Banerjee et al. 2007; Duflo et al. 2009; Muralidharan and Sundararaman 2010; all cited in Bruns et al.). The authors point out the fact that contracted positions often incorporate alternative curricula, have different class makeups, and are seen as temporary posts while teachers wait for more permanent civil service appointments.

<sup>11</sup> Muralidharan and Sundararaman (2009; cited in Bruns et al.). The study incorporated 500 state schools in Andhra Pradesh randomly assigned to either a group or individual pay-for-performance scheme, contract teachers, institutional block grants, or a control group. Findings were robust after the second year of administration (0.27 SD improvement compared to control).

to test various hypotheses in social context. In their chapter on learning in LICs, the authors elaborate on a major challenge—how to maintain strong school attendance in both students and teachers. To illustrate the challenge of ensuring that teachers show up for a full day of instruction, the authors describe service delivery breakdowns resulting from teacher protests—a common occurrence in LICs due to teacher strikes. The negative effects are passed along to the schoolchildren who suffer from shortened learning opportunities. To offset teacher absenteeism, some countries have begun experimenting with innovative accountability interventions.

In Rajasthan (India), for example, Karlan and Appel describe a program that reduced teacher absenteeism by half through the distribution of disposable cameras to school children for documenting and linking teacher attendance to a pay-per-performance incentive scale (214).<sup>12</sup> However, as the authors point out, monitoring tools need to be “corruption-proof” for changes to be effective (231). For instance, an adaptation of the Rajasthan study linked health clinic workers’ salaries to attendance schedules and resulted in inconclusive findings. The authors describe broken time-stamping machines and supervisor exemption authority (ostensibly to allow for routine offsite duties) as potential limiting factors (230). Again, the Rajasthan studies point to the significance of context within development work and the importance of a well-functioning institution for carrying out program objectives.

Each volume invokes substantial empirical research when discussing findings and implications, often providing important insights into development work. Nonetheless, the ability to extrapolate from controlled evidence to predict real-world impacts raises serious issues, particularly in situations with highly varied contexts.

### **Strategic Perspectives on Social Experiments**

Together, these three volumes share the broad view that social experiments—mainly in the form of EB evaluations using RCT designs—offer new and more credible ways to confirm (or disconfirm) hypotheses about cause and effect in international education development. Together, they also suggest a number of strategic perspectives that educational specialists might consider when making future investments.

#### *Focus on Human Agency*

Banerjee and Duflo contend that “the ladders to get out of the poverty trap exist but are not always in the right place, and people do not seem to know how to step onto them or even want to do so” (50). In the area of health, their findings showed that families often spent (counterproductively)

<sup>12</sup> Duflo et al. (2010); cited in both Karlan and Appel and Bruns et al. Additionally, the study found a major increase (about 33 percent) in time on task—which then translated into higher test scores among treatment schools.

more on expensive cures rather than on cheaper preventative care. In education, they found that parents usually underestimated the value of more schooling for their children, especially girls. Banerjee and Duflo show compelling evidence of the rationality of decision making of the poor and how a focus on changing behavior—building on human agency—can have significant impacts on children and education.

#### *Nudges for Development*

Both Banerjee and Duflo as well as Karlan and Appel emphasize the importance of carefully designed “nudges” as a means of promoting behavioral change. Banerjee and Duflo emphasize the importance of convincing individuals of the benefit of a behavior change rather than trying to “bribe them to do what *you* think is right” (63). For example, they describe the apparent inability of many poor people to move from intention to action—namely, that humans tend to perceive the future very differently than the way they see the present (64). Karlan and Appel point to the success that they observed through applying this approach for microfinance by sending low-cost blast text messages and e-mails to bank clients to remind them to save (164). Human nature influences how we respond to immediate inconveniences when it comes to the domains of health (e.g., deciding not to stand in line to receive the final cycle of an immunization) as well as education (e.g., child-labor income taken now as opposed to investing in school years to boost income in the future). In development, even small nudges can have a major impact.<sup>13</sup>

#### *Inverse Incentives*

Contrary to nudges that promote an action against the status quo, inverse incentives are designed to make it harder to opt out of a particular behavior of interest. Related to this notion, Banerjee and Duflo discuss the important concept of designing interventions with a default behavior embedded into them that is regarded as the desired outcome (65). In other words, individuals may move away from a (developmentally good) action—but at a real cost. In a similar way, Karlan and Appel discuss the benefit of commitment contracts. These incorporate user-authorized penalties to discourage the user’s procrastination toward a desired goal—such as a savings plan, weight loss, or smoking cessation (159).

#### *Rigor and Transparency*

As the body of research using social experiments continues to grow, systematic reviews and meta-analyses, as in Bruns et al., will be deployed for building more robust conclusions above that of considering individual studies

<sup>13</sup> For examples of intervention nudges, see Thaler and Sunstein (2008; cited in Banerjee and Duflo).

separately.<sup>14</sup> Additional efforts will no doubt improve the transparency of research methodologies and findings across the diverse contexts of international development. In the three books reviewed here, the authors emphasize rigorous methods as key to EB evaluations and more credible results. The temptation, of course, is for development agencies and policy makers to view other approaches (e.g., qualitative, ethnographic, and mixed methods) as necessarily lacking in rigor and credibility. Solid arguments in these books and elsewhere argue against this simplistic view.<sup>15</sup>

### Conclusion

Researchers and policy makers are increasingly faced with the task of making sense of a complex and variegated landscape of empirical outcomes. Attempting to collate and integrate findings from one methodology to another, and across an increasingly diverse set of social and cultural contexts, is daunting. A salient theme from this review is that evidence in international development—whether EB, RCT, or other methodological approaches—is neither gathered in a vacuum nor in a well-controlled laboratory. Contextual variation is difficult to control, and findings in diverse development contexts contain significant barriers to generalization. These volumes demonstrate that effective interventions can have a profound impact on educational outcomes. But impact, scale-up, and sustainability are undermined when the cultural dimensions of human behavior are insufficiently understood and institutional accountability is weak. Let the buyer of the gold standard beware.

### References

- Aikman, S. 1999. "Schooling and Development in the Amazon." In *Gender, Education, and Development*, ed. C. Heward and S. Bunwaree. New York: Zed.
- Baird, S., C. McIntosh, and B. Ozler. 2009. "Designing Cost-Effective Cash Transfer Programs to Boost Schooling among Young Women in Sub-Saharan Africa." World Bank Policy Research Working Paper no. 5090, World Bank, Washington, DC.
- Banerjee, Abhijit V., S. Cole, Esther Duflo, and L. Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *Quarterly Journal of Economics* 122 (3): 1235–64.
- Benhassine, N., F. Devoto, E. Duflo, P. Dupas, and V. Pouliquen. 2010. "The Impact of Conditional Cash Transfers on Schooling and Learning: Preliminary Evidence from the Tayssir Pilot in Morocco." Mimeo, Massachusetts Institute of Technology, Cambridge, MA.

<sup>14</sup> See Duvendack et al. (2012).

<sup>15</sup> For example, the use of mixed-methods approaches that incorporate qualitative modes of inquiry to complement RCT designs and quasi-experimental research are providing a more comprehensive perspective for overcoming methodological divides in the realm of education development (e.g., Burde 2012). While assessing impact and establishing causal inference are critical, understanding processes shaped by context and implementation must also be considered (Mallet et al. 2012).

- Burde, D. 2012. "Assessing Impact and Bridging Methodological Divides: Randomized Controlled Trials in Countries Affected by Conflict." *Comparative Education Review* 56 (3): 448–73.
- Clammer, J. 2012. *Culture, Development and Social Theory: Towards an Integrated Social Development*. London: Zed.
- Duflo, Esther, P. Dupas, and M. Kremer. 2009. "Additional Resources versus Organizational Changes in Education: Experimental Evidence from Kenya." Unpublished manuscript, Abdul Latif Jameel Poverty Action Lab (JPAL), Massachusetts Institute of Technology, Cambridge, MA.
- Duflo, E., R. Hanna, and S. Ryan. 2010. "Incentives Work: Getting Teachers to Come to School." Unpublished manuscript, Abdul Latif Jameel Poverty Action Lab (JPAL), Massachusetts Institute of Technology, Cambridge, MA.
- Duvendack, M., J. Garcia Hombrados, and R. Palmer-Jones. 2012. "Assessing 'What Works' in International Development: Meta-Analysis for Sophisticated Dummies." *Journal of Development Effectiveness* 4 (3): 456–71.
- Farr, S. 2010. *Teaching as Leadership: The Highly Effective Teacher's Guide to Closing the Achievement Gap*. San Francisco: Jossey-Bass.
- Fiszbein, A., and N. Schady. 2009. *Conditional Cash Transfers: Reducing Present and Future Poverty*. Washington, DC: World Bank.
- Glewwe, P., M. Kremer, and S. Moulin. 2000. "Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya." Bureau for Research and Economic Analysis of Development working paper, Duke University. <http://ipl.econ.duke.edu/bread/index.htm>.
- Hanushek, E. A., and S. G. Rivkin. 2010. "Generalizations about Using Value-Added Measures of Teacher Quality." *American Economic Review* 100 (2): 267–71.
- Lomeli, E. 2008. "Conditional Cash Transfers as Social Policy in Latin America: An Assessment of Their Contributions and Limitations." *Annual Review of Sociology* 34 (1): 475–99.
- Mallet, R., J. Hagen-Zanker, R. Slater, and M. Duvendack. 2012. "The Benefits and Challenges of Using Systematic Reviews in International Development Research." *Journal of Development Effectiveness* 4 (3): 445–55.
- Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72 (1): 159–217.
- Muralidharan, K., and V. Sundararaman. 2009. "Teacher Performance Pay: Experimental Evidence from India." NBER Working Paper no. 15323, National Bureau of Economic Research, Cambridge, MA.
- Muralidharan, K., and V. Sundararaman. 2010. "Contract Teachers: Experimental Evidence from India." Unpublished paper, World Bank, Washington, DC.
- Ranganathan, M., and M. Lagarde. 2012. "Promoting Healthy Behaviours and Improving Health Outcomes in Low and Middle Income Countries: A Review of the Impact of Conditional Cash Transfer Programmes." Supplement, *Preventive Medicine* 55:S95–S105. doi: 10.1016/j.ypmed.2011.11.015.
- Thaler, R. H., C. R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New York: Penguin.
- Wagner, D. A. 2011. *Smaller, Quicker, Cheaper: Improving Learning Assessments for Developing Countries*. Paris: UNESCO-IIEP.