

Science of delivery: Education outcomes and the World Bank

Arianne Wessal, Clay Wescott and Carlos Espindola

This article addresses the following research question: do World Bank-supported education projects achieve better outcomes when there is deeper attention to designing monitoring and evaluation systems during project design?

The work presented here draws on a recently constructed database of World Bank investment projects. Ratings of education projects' monitoring and evaluation design, and their outcomes, will be compared to see whether there is a correlation between high monitoring and evaluation design ratings, and better outcomes. This pattern has been observed in other service areas supported by the World Bank. Further analysis will look at other factors associated with better outcomes and determine their relative importance. The next step will be to analyse detailed project components to see if there are design features common to high performing projects with effective monitoring and evaluation, and other outcome-enabling features that are missing in weaker performing projects. Further testing will be carried out using selected case studies.

Evaluations of World Bank projects have, in recent years, shown a decline in outcome ratings. Improvements in monitoring and evaluation, and other features, may point the way to reversing this trend. When the enabling environment encourages experimentation, tight feedback mechanisms and constant communication make it possible for project managers to make real-time changes to projects throughout the project cycle. Feedback mechanisms could include such measures as evaluation committees, seminars and workshops, automated systems, reporting and follow-up procedures. It is also essential to tailor project components to local factors, such as implementation capacity and political support. Project managers could be given greater incentives to draw on aspects of past successful projects, try new concepts and adapt to changing conditions.

Research on achieving results

Many developing country governments are trying to understand why the policies put in place to reduce poverty and build prosperity are not leading to the results they want. One way forward could be a new form of knowledge, the 'science of delivery'. This concept is borrowed from the health care field, where the previous emphasis on understanding the causes and consequences of health issues is

shifting to give more attention to organising, managing and financing health promotion (Catford, 2009). Applied to the field of public management, the science of delivery should provide mechanism-based explanations of how and why the implementation capability of countries varies, as well as a guide to action (Woolcock, 2013). This approach differs from the institutional reform model that currently dominates the public management field. In the institutional reform model, best practice 'solutions are often chosen without significant consideration being given to their external validity. In this model, the focus is on inputs delivered rather than on outputs obtained. Another pitfall is that reforms try to take on too much and are stymied by the complexity, lack of clarity, uncertainty and unintended consequences of efforts to change socio-political structures' (Schuck, 2014: p. 372). The result is that projects frequently fail to achieve their goals, while the specific reasons for this failure remain hard to pin down.

In order to remedy these issues, the science of delivery tailors project components based on local factors, such as implementation capacity and political support. As problems arise, consideration is given to concerns at the political, organisational and project levels before deciding on a solution. Project managers are encouraged to draw on processes linked to successful projects, try new concepts and adapt to changing conditions. The science of delivery approach requires experimentation, intensive field research documented in accessible case studies, improved data collection at the project level through the use of good monitoring systems, and the diffusion of ideas to enable these changes in implementation and management. The result of using a science of delivery approach is the incremental creation of localised projects that provide impactful results to the target community as well as useful data and information to the public. This data gives project managers the ability to understand how and why a project was effective rather than just whether it was or not. The science of delivery allows project managers in a region to better understand why their projects fail to achieve their desired impact, as well as giving them the ability to draw on lessons learned from successful projects in other regions. At the same time, there have been recent theoretical advances in many scholarly fields including systems engineering, medicine, economics and public management, which are being exploited to help countries organise the emerging evidence on successful delivery to help them improve development results (Kim,

This paper has not undergone the review accorded to official World Bank publications. The findings, interpretations and conclusions expressed herein are those of the author(s) and do not necessarily reflect the views of the International Bank for Reconstruction and Development, the World Bank and its affiliated organisations, or those of the executive directors of the World Bank or the governments they represent. The World Bank does not guarantee the accuracy of the data included in this work. This paper draws on Wessal, A., Treuth, M. and Wescott, C., 2014, Science of Delivery and Implications for Monitoring and Evaluation. An expanded version was presented at City University, Hong Kong, in December 2014 at an International Public Management Network conference, 'The Science of Delivery'. The authors are grateful for a peer review by Alison Wescott.

2012). These new sources of knowledge help aid managers in adapting their projects to local conditions, ultimately resulting in a higher level of success.

The World Bank and other development partners can point to many examples of delivery success, drawing on a treasure trove of evidence obtained using a mix of qualitative and quantitative methods, linking successful delivery of interventions with local politics, culture, capacity and other factors. However, some of this experience is not easily accessible and may be buried in lengthy reports, files, datasets, and the heads of staff and evaluators as tacit knowledge.

Education

How can a science of delivery approach improve schooling? Education is a critical public service and research in developing countries reconfirms that education can enable growth and poverty reduction (Schultz, 1993; Bloom et al., 2006). However, this causal chain is not always straightforward: for example, learning outcomes of poor children may be held back by health, nutrition and social factors during their first five years of age, such as stunting, inadequate cognitive stimulation, iodine deficiency and iron deficiency anaemia (Walker et al., 2007). Indeed, one cross-national study found no link between rising education of the

Project examples

Further analysis will be conducted to see whether there are specific design features that are common to the high-performing projects with effective monitoring and evaluation. Some possible features are evident from looking at three examples where strong monitoring systems have been built at the beginning of the programmes with short-, medium- and long-term outcomes identified.

Jamaica: PATH

In the case of DfID's Programme of Advancement through Health and Education (PATH), process evaluations and spot checks were undertaken for activities being implemented, allowing for the identification of any potential issues (Rawlings, 2009). As a result of these evaluations, the application process was deemed to be burdensome and unclear on programme rules, the system for verifying the eligibility of new beneficiaries was weak, and there was found to be a strong unmet demand for jobs and training. Quality data enabled management to exploit double-loop learning (Argyris and Schön, 1974), which led to a decision to revamp the management information system, revise the operations manual, use social workers as focal points to access social services, and create a 'STEPS to Work' programme focused on skill development and employment. Evaluations of PATH showed that it was better at reaching the poor than other Jamaican safety net programmes.

Mexico: Oportunidades (formerly Progresa)

Evaluations of Oportunidades showed the programme had a significant positive impact in improving health and education, and it has been lauded for reaching its target populations and yielding better results than other programmes. Oportunidades is a great example of improved science of delivery through the use of both a strong monitoring and evaluation system and of information learned from past projects. The programme began in 1997, providing monetary educational grants to poor, rural families for each child enrolled in school between the third grade of primary and third grade of high school. In addition to education, Oportunidades also has health and nutrition components. Government health institutions provide families with preventative health care and families also receive a fixed monthly transfer to improve food consumption. Nutritional supplements are provided for young children and their mothers.

Where Oportunidades truly shines is in quality at entry. At implementation, project managers planned to have an

independent evaluation done by the International Food Policy Research Institute. This independent evaluation was planned with the goal of providing management with data to make real-time project changes and improve delivery. They also drew on lessons learned from past projects, recognising that giving money to female heads of families results in better financial outcomes. These steps, taken during the implementation stages, translated into quality results that were reflected in the independent evaluation – the first large-scale, randomised controlled trial used in developing social policy. The results caught the eye of the Mexican federal government, and an increase in funding allowed Oportunidades (now accounting for 46.5 per cent of the federal anti-poverty budget) to expand to urban areas and to provide high school students with education grants. In summary, steps taken at implementation to improve the science of delivery were crucial in the success and subsequent expansion. The close involvement of scholar-practitioners helped to design new conceptual approaches, ensure technical soundness and rigorous monitoring, protect the programme during changes of administration, and spread the approach around the world (Lustig, 2011).

Brazil: Minas Gerais Development Partnership

In 2008 a sector-wide approach of more than US\$1.4 billion was initiated, aimed at improving the efficiency of public resource use, supporting innovations in public management, and supporting the state government of Minas Gerais in strengthening its monitoring and evaluation system (World Bank, 2008). Funds were disbursed to ten eligible expenditure programmes in five sectors. Individual projects subject to monthly monitoring and quarterly management meetings were made accessible to the press. The government made yearly implementation data available on the web to increase programme transparency and improve data dissemination. The World Bank supported the project by developing a household survey, quality assurance surveys and a series of impact evaluations in the education, health and transport sectors. These monitoring systems provided managers necessary feedback mechanisms and learning loops, allowing them to work towards the achievement of medium-term goals on their way to the achievement of long-term objectives. So far, the programme has succeeded in reducing the amount of time needed to start a business at Minas Facil in Belo Horizonte from 26 to seven days. The poverty reduction programme has already exceeded its initial objective by benefiting more than 26,000 rural families.

labour force and growth in output per worker, perhaps due to a combination of perverse governance environments, stagnant demand for educated labor and poor educational quality (Pritchett, 2001). Another cross-national study focused on cognitive skills measured by international tests, rather than school attainment, and found a strong link to economic growth independent of other contributing factors, such as property rights, open markets and effective economic institutions (Hanushek and Woessmann, 2008). Thus, attaining societal benefits depends on people acquiring knowledge and skills, not just on attending school (Boissiere, 2004; Glewwe, 2002).¹

The choice of a particular focus of school reform differs as an education system moves from one performance level to the next. Moving from poor to fair performance, for example, typically has a focus on gathering information. It should be noted that in some regions schools share school performance data publicly, while

others share such data only privately among schools (Mourshed et al., 2010; Banerjee and Duflo, 2011: pp. 97–101).

World Bank investment projects

Comparing the World Bank’s investment projects completed between the 2007–09 financial years with those completed in 2010–12, the number of projects rated high or substantial on monitoring and evaluation declined from 33 to 26 per cent. Reasons included limited baseline data, unfocused indicators and indicators measuring just outputs rather than outcomes. There were also weak monitoring organisations with high staff turnover and vacancies, unclear roles and responsibilities for data collection, weak management information systems and data quality, and limited data utilisation for decision making (IEG, 2014: pp. 45). This weakening monitoring and evaluation performance has contributed to an overall decline in outcomes across all sectors.

Figure 1: Project ratings

Outcome rating	Quality ratings		
	Modest and above	Negligible	Total
Moderately satisfactory and above	50	1	51
Moderately unsatisfactory and below	20	5	25
Total	70	6	76

Figure 2: Project ratings

Outcome rating	Quality ratings		
	Modest	Substantial	Total
Moderately satisfactory	26	6	32
Satisfactory	7	11	18
Total	33	17	50

Figure 3: Statistical significance

Outcome rating	Quality ratings		
	Modest and above	Negligible	Total
Moderately satisfactory and above	50 (46.97) [0.19]	1 (4.03) [2.27]	51
Moderately unsatisfactory and below	20 (23.03) [0.4]	5 (1.97) [4.64]	25
Total	70	6	76

The figures in parentheses represent, respectively, the expected cell totals if the variables were independent and the chi-square statistics for each cell

Figure 4: Ratings by region

	Africa	East Asia and the Pacific	Europe and Central Asia	Latin America and the Caribbean	Middle East and North Africa	South Asia
Modest and above	89%	100%	100%	89%	73%	92%
Negligible	5%	0%	0%	11%	18%	8%
Total	94%	100%	100%	100%	91%	100%

The same ratings are also declining in the education sector and have been since financial year 2001. In addition to the decline in monitoring and evaluation quality, there were many other factors judged to be more important, about half (seven of the 15) concerning design and half implementation/supervision. These included design issues such as: over ambition in relation to the strength of political commitment; over ambition in relation to the time period; inadequate readiness for implementation; weaknesses in technical design, including prior analytical work; and over ambition or excessive complexity with regard to a country's institutional or implementing capacity.

Analysis

Using the recently constructed Independent Evaluation Group (IEG) database of World Bank projects and their components, an initial analysis of the projects within the database was taken in order to see whether there is a correlation between high monitoring and evaluation ratings and high outcome ratings in World Bank education projects, as validated in the implementation completion report reviews.²

The results were compared using a two-by-two table in which projects with IEG outcome ratings that were moderately satisfactory³ and above were grouped, and those that were moderately unsatisfactory and below were grouped. These were compared to monitoring and evaluation quality ratings (quality ratings) with categories of modest⁴ and above, and the category of negligible. In this case there were a total of 63 projects with both IEG outcome ratings, and monitoring and evaluation ratings.

As seen in Figure 1, the analysis showed that the majority of the projects had quality ratings of modest and above, and outcome ratings that were moderately satisfactory or above, while less than half of projects with a modest and above monitoring and evaluation rating had moderately unsatisfactory and below outcome ratings.

A breakdown of the 41 projects with moderately satisfactory and satisfactory outcome ratings, and modest and substantial quality ratings, shows that only ten projects had an outcome rating of satisfactory and a monitoring and evaluation rating of substantial, as shown in Figure 2. Although the highest monitoring and evaluation rating possible is ranked high, there were no projects with a high monitoring and evaluation rating in the 63 analysed.

To test the statistical significance of the correlation between the quality rating and the IEG outcome rating, a chi-squared test was run and found that the results were statistically significant at the less than 95 per cent level (see Figure 3). A chi-square test is a commonly used statistical test which tests the correlation between two variables of interest. The test looks at whether deviations in the data from theoretical proportions have occurred by chance or not. For that purpose the test uses the chi-square distribution to estimate the theoretical proportions of how the data should behave.

In addition, a breakdown of the monitoring and evaluation ratings by regions for this data set showed East Asia and the Pacific and Europe and Central Asia to have the highest ratings (see Figure 4).

Of the 41 projects with moderately satisfactory and above outcome ratings, and modest and above quality ratings, 95 per cent of the component records associated with these had monitoring and

evaluation as a component or mentioned monitoring and evaluation explicitly as part of their component descriptions.

Conclusions

The results that emerged from using a recently constructed database of World Bank projects and their components, and linking to IEG ratings, show that better performing projects do have higher monitoring and evaluation ratings, and that the difference is statistically significant. However, even the better performing projects mainly have only a modest rating for monitoring and evaluation, suggesting room for improvement. Initial case study analysis found specific design features that are common to the high performing projects with effective monitoring and evaluation. There are also added benefits when projects are designed, encouraging experimentation and taking into account distinctive features of the local context. This analysis needs to be extended to the rest of the sample to see if these features are common, and if so, the extent to which they help to explain the improved performance.

Endnotes

- 1 Impact evaluations point to three approaches that are most effective in improving learning outcomes. First of all, information reforms, such as providing school and student test scores, help to make better comparisons among schools, improve student performance and reduce school fees (Andrabi et al., 2014). Second, school-based management reforms, such as tracking students based on prior achievement into separate classes, increasing school autonomy and empowering parents can be effective, although there are only a small number of rigorous studies and the metrics vary in different studies. Third, teacher incentive reforms can have an impact, making teachers more accountable for results by linking tenure and/or pay to performance. Again, the evidence base on this is small, but promising (Bruns et al., 2011; Duflo et al., 2008).
- 2 The analysis was conducted by using all the education projects with monitoring and evaluation quality ratings, and outcome ratings that we found in both the IEG components database (privately available) and in IEG's publicly available database of World Bank project ratings. Only projects with both monitoring and evaluation quality ratings and IEG outcome ratings that overlapped in both databases were used for the analysis. There were a total of 63 projects that fell into this category. Only this population of projects was used in the analysis to ensure that other researchers could replicate this exercise if desired. The completion dates for the implementation completion report reviews all fall between 2007 and 2012.
- 3 Outcome ratings are on a six-point scale: high, satisfactory, moderately satisfactory, moderately unsatisfactory, unsatisfactory and highly unsatisfactory.
- 4 Monitoring and evaluation ratings are on a four-point scale: high, substantial, modest and negligible.

References

- Andrabi, T., Das, J. and Khwaja, A. I., 2014. *Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets* [pdf] Harvard Kennedy School, Harvard. Available at: www.hks.harvard.edu/fs/akhwaja/papers/RC_June2014.pdf [Accessed 1 December 2014].
- Argyris, C. and Schön, D., 1974. *Theory in Practice: Increasing Professional Effectiveness*. San Francisco: Jossey-Bass. pp. 2–3.
- Banerjee, A. and Duflo, E., 2011. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. New York: Public Affairs.

Bloom, D., Hartley, M. and Rosovsky, H., 2006. 'Beyond private gain: The public benefits of higher education'. In: Samuel Neaman Institute, 2004. *Transition to Mass Higher Education Systems: International Comparisons and Perspectives*. Haifa: Samuel Neaman Press.

Boissiere, M., 2004. *Rationale for Public Investment in Primary Education in Developing Countries* [pdf] IEG (Independent Evaluation Group). Available at: http://ieg.worldbank.org/Data/reports/education_public_investments_wp.pdf [Accessed 1 December 2014].

ARIANNE WESSAL (awessal@worldbank.org) is a monitoring and evaluation consultant with the Global Partnership for Education, which is administered by the World Bank. Prior to this, she worked with the World Bank's Independent Evaluation Group (IEG), with a focus on capacity-building and global programmes. She has also worked on education and social development with Save the Children, the Inter-American Development Bank and the World Bank. Her research has focused on issues of access and equity to basic education in rural areas with particular attention to nomadic populations. Wessal holds a master's degree in international development policy from Duke University.

CLAY WESCOTT is director of the Asia-Pacific Governance Institute, president of the International Public Management Network and a senior consultant with the World Bank. He is also the book review editor for *Governance*, and senior editor of the *International Public Management Review*. He has held senior positions with Princeton and Harvard universities, the Asian Development Bank, United Nations Development Project (UNDP) and Development Alternatives Inc. He has degrees from the universities of Harvard and Boston.

CARLOS ESPINDOLA is a managing associate of the Infrastructure Planning and Economics group at Nathan Associates. He is a registered transport engineer and holds an MBA focused on strategic and regional planning for transport and logistics projects. He is frequently involved in project design and performance assessment as part of his work, much of which has focused on Latin America, the USA and Africa. Espindola also works with engineering and economic consulting firms on projects sponsored by diverse clients, including multilateral developing partners, US foreign aid agencies, national governments and private investors.

Bruns, B., Filmer, D. and Patrinos, H., 2011. *Making Schools Work: New Evidence on Accountability Reforms*. Washington: World Bank.

Catford, J., 2013. 'Advancing the "science of delivery" of health promotion: Not just the "science of discovery"'. *Health Promotion International*, 24 (1), pp. 1–5.

Duflo, E., Dupas, P. and Kremer, M., 2008. 'Peer effects, teacher incentives, and the impact of tracking: evidence from a randomized evaluation in Kenya'. *American Economic Review*, 101 (5), pp. 1739–1774.

Glewwe, P., 2002. 'Schools and skills in developing countries: Education policies and socioeconomic outcomes'. *Journal of Economic Literature*, 40 (2), pp. 436–82.

IEG (Independent Evaluation Group), 2014. *Results and Performance of the World Bank Group 2013* [webpage] World Bank. Available at: <http://ieg.worldbankgroup.org/evaluations/rap2013> [Accessed 2 December 2014].

Lustig, N. C., 2011. *Scholars Who Became Practitioners: The Influence of Research on the Design, Evaluation and Political Survival of Mexico's Antipoverty Program Progresar/Oportunidades* [webpage] Tulane University. Available at: http://stonecenter.tulane.edu/uploads/Lustig_Scholars_and_Practitioners_Mx_June_2011_correct_version-1308242989.pdf [Accessed 2 December 2014].

Mourshed, M., Chijioke, C. and Barber, M., 2010. *How the World's Most Improved School Systems Keep Getting Better* [webpage] McKinsey and Company. Available at: www.mckinsey.com/client_service/social_sector/latest_thinking/worlds_most_improved_schools [Accessed 2 December 2014].

Rawlings, L., 2009. *Monitoring and Evaluation: The Foundations for Results* [pdf] World Bank. Available at: http://siteresources.worldbank.org/INTISPMA/Resources/383704-1184250322738/3986044-1250881992889/M&E_TheFoundationsOfResults_Rawlings_EN.pdf [Accessed 2 December 2014].

Schuck, P., 2014. *Why Government Fails so Often and How it can do Better?* Princeton: Princeton University Press.

Schultz, T. P., 1993. 'Returns to Women's Education'. In: E. M. King and M. A. Hill, eds. 1997. *Women's Education in Developing Countries: Barriers, Benefits and Policies*. Baltimore: Johns Hopkins University, pp. 51–99.

Walker, S. et al., 2007. 'Child development: Risk factors for adverse outcomes in developing countries'. *The Lancet*, 369 (9556), pp. 145–57.

Woolcock, M., 2013. *The Science Of Delivery-Taking Context Seriously: Expanding the Range of Ideas, Methods and Strategies to Enhance Development Effectiveness* [video] World Bank Group. Available at: https://cdnapisec.kaltura.com/index.php/extwidget/openGraph/wid/1_pegt1z2f [Accessed 2 December 2014].