**Smart Buys, Great Sales and Special Offers**

**Is the Global Education Evidence Advisory Panel Report a Good Buy?**

**Keith M Lewin,**

**Emeritus Professor of International Development and Education**

**University of Sussex**

**Draft 0 Version[[1]](#footnote-1)**

**Preamble**

The Global Education Evidence Advisory Panel (GEEAP) launched its report on “Cost Effective Approaches to Improve Global Learning Levels: What does recent evidence tell us about “Smart Buys” for improving learning in low and low middle income countries?” at the end of October. Learning and the opportunity to acquire and apply acknowledge and skill has always been at the heart of development. Any efforts that increase cost effectiveness and make it more likely that countries can manage and develop their education systems from domestic resources without repeated cycles of aid must be welcomed.

It is more than 50 years since Phil Coombs wrote the “World Education Crisis” and thirty years since the World Conference on Education for All launched the global Framework for Action to Meet Global Learning Needs in 1990. This report should stand on the shoulders of the many previous collections that speak to the question of how to improve learning including various editions of the Global Education Monitoring Report (GEMR), the McKinsey Reports on the world’s best performing schools systems, the World 2018 Development Report of the World Bank on education, the Global Partnership for Education’s Knowledge and Information Exchange (KIX), Research on Improving Systems of Education (RISE), and Progamme for International Student Assessment (PISA) to name a few of the more prominent ones. Readers will want to compare the findings of the GEEAP report with those of many other independent analyst of education and development to see how they resonate. They will also be interested to understand how this initiative maps onto the existing architecture of aid to education. Are these best buys already part of these plans or have they been over-looked? Do they now have to be grafted on and what is the appetite from governments to act on good and bad buys?

The [Global Education Evidence Advisory Panel (GEEAP)](https://www.worldbank.org/en/topic/teachingandlearning/brief/global-education-evidence-advisory-panel) has launched its report on ‘[Cost Effective Approaches to Improve Global Learning](https://www.worldbank.org/en/topic/teachingandlearning/publication/cost-effective-approaches-to-improve-global-learning)’. Readers of the GEEAP report will want to compare and cross reference its findings on achievement with those already in the public domain and widely discussed at UKFIET Conferences in the past, to see how the recommendations resonate. They will also be interested to understand how this initiative maps onto the architecture of aid to education, which  should continue to develop. Are these best buys already part of this architecture or have they been persistently over-looked? Do they now have to be grafted on to existing projects? What is the appetite from governments to adopt these good and bad buys. Will they continue to buy with their own money after the external finance stops? The report presents a taxonomy of interventions that are grouped under five headings.



The GEEAP report identifies one ‘great buy’, six ‘good buys’, three ‘promising options’ with “low” evidence, three ‘bad buys’, and five “Areas where governments need to make decisions or take action”. The last category includes investment in general in-service teacher training, the selection and deployment of teachers, and interventions related to gender, disability and violence though the evidence base for these is described as “low”.



Readers will make their own judgements as to how comprehensive and novel the lists are and whether the sampling and case studies can support global generalisations. In most of the categories there are studies that show significant effects and other studies that indicate weaker or no impact of interventions. Averages may not be the best way to proceed since they are applied to support generalisations that only apply to cases close to the central tendency.

**Buying or Selling?**

An apocryphal story poses a central dilemma relevant to Best buys.

“Margaret Thatcher was asked about the price of eggs by an aspirant voter. The problem posed was that chickens lay on average 0.7 eggs each day, they eat 1.5 pounds of feed a week and feed is about £0.75 pence per pound. So how much is the price of an egg? She turned, looked quizzically into her famous handbag and said with a glint in her eye “buying or selling?”

Key issues for the GEEAP are: who is buying what, from whom, who is the Principal and who is the Agent in the transactions? What are governments, NGOs and donors buying and selling? Is there really a market that determines price, or is it a simply a framework for contracting services? Is a ‘smart buy’ the one that is best for the seller or for the buyer? If the best buy is also the best sale (e.g. performance related development bonds?), is everyone in the market ‘smart’? Are smart buys those where costs are very low relative to small benefits, or where benefits are very large relative to substantial costs? Which bargains are value for money but unaffordable? The question is how to avoid the dissonance that arises from seeking best buys for discrete educational interventions that are not part of a system specific development plan. If aid is to be directed towards the generic best buys as is planned, then it will lack local coherence with independently developed national plans.

The question is how to avoid the *collection and consumption trap* that arises from seeking best buys for discrete educational interventions. What if I shopped at Tesco/Walmart without a shopping list and simply bought those things that had the best special offers and which offered the greatest satisfaction at lowest cost. What if “pick and mix” was the order of the day in the supermarket with the price determined by the weight rather than the quality of the sweets? I would leave the store happy but my bargain basket would most likely contain own brand chocolate, fizzy drinks and value packs of crisps and a long term increased risk of obesity. When I arrived home I might also find the fridge was already stocked with great buys from my previous shopping expedition!

Surely lists of best buys in development have to start from a clear set of system level goals and metrics that link educational investment to diagnosis of development needs that are whole sectoral in scope. Investment at one level interacts with other levels. The buys have to fit into sector wide plans designed to respond to development needs that are broadly defined. Learning is much more than literacy and numeracy.

**Caveat Emptor**

The GEEAP report identifies one Great buy, six Good buys, three Promising Options with “low” evidence, three Bad buys, and five “Areas where governments need to make decisions or take action” also with little evidence on impact. It is built from high quality evaluation studies of externally financed interventions. It deploys Randomised Control Trials (RCTs) that can be good at identifying effects of inputs on outputs under tightly controlled conditions and in static systems. RCTs are not so well suited to interventions that are multi-dimensional, interactive and rapidly evolving. They can give causal insight into closed systems but are not as helpful in understanding open systems like education or development assistance These are better explored using other research methods grounded in organisational context and process, and theories of innovation and the management of planned change.

Readers will make their own judgements as to how comprehensive and novel are the lists of good and bad buys. Critically, what you see is all there is (WYSIATI) does not apply[[2]](#footnote-2). The choice of buys is restricted to those interventions that are visible because they have been externally funded and have had RCT based evaluations. There is therefore a confirmation bias in this sample towards existing interventions that excludes many others that can be imagined but which have been undertaken without aid and which cannot be evaluated by an RCT.

Surprisingly the report only identifies one Great buy after more than 50 years of large scale development assistance, several World conferences, and thousands of well-considered academic papers and project evaluations. The Great buy of “Giving information on education benefits, costs and quality” (p10) appears to be based on a sample of only two country case studies (Figure 1, p32, N=2)). These studies have a very high average cost benefit because one (Mozambique) has a very high value whilst the other registers almost no effect. Averages have limited uses in these circumstances. The cost benefit is also so high because the costs are so small.

There will also be puzzlement that only three Bad buys are identified despite the persistent global crisis in access and learning that must have something to do with past choices. These bad buys include many of the things that Ministries currently invest in including “textbooks, technology hardware, uniforms, school grants, class size reductions, and general skill teacher training”. The analysis indicates on the margin these have little or no effect on learning with the implication that money would be better spent on other things such as “early childhood development and merit based scholarships for the disadvantaged”. This is a difficult conclusion to sell. It is counter intuitive since, for example, there are upper limits of class size and consequences to having no text books and ignores the political economy of choice.

These and other limitations need open discussion. First, it seems many of the studies do not have cost data on direct and indirect costs, opportunity costs and fully inclusive donor costs. About “half of all interventions (that have cost data) produce no significant effects: these interventions are not included in the figures”. Why they are omitted? Knowing which interventions have no effect and why would seem as important at knowing which ones do have an effect.

Second, the buying classifications are difficult to understand. How were the boundaries decided between Great, Good, Promising and Bad since they do not simply follow the average score on Learning Adjusted Years of Schooling (LAYs). Did the process review historical experience with each type of innovation? The lists provided in Annex D indicate that of the eleven types of interventions listed seven have low cost effectiveness between zero and 4 LAYs with a wide intra category variation in effectiveness. Three had no impact at all. Cherry picking those that have impact and ignoring those that have no impact cannot not tell us why interventions fail.

Third, GEEAP asserts that “the classifications and descriptions in this note aim to offer a greater sense of prioritisation and clarity than has been possible in the past”. If so then how can this be translated into educational plans related to the SDGs and loan programmes and implementation grants for development assistance? What exactly are the new and previously invisible prioritisations not already anticipated in past and present projects? Which current priorities should be downgraded to make fiscal space?

Fourth, the number of cases for cost effectiveness appears too small to lead to robust conclusions. Four of the seven Great/Best buys have fewer than 4 country case studies (Figure 1) each of which includes cases where the effect is close to zero.

Fifth, the insights that inform the GEEAP taxonomy are from RCTs that can provide useful within the constraints of their design. But surely the Best buys identified are not new to those familiar with education and development over the last 50 years. They also have precedents in many project planning documents and country analyses of the World Bank and bi-lateral donors. The GEEAP needs a much clearer vision of who it is providing advice to and what is it unique contribution over and above all the parallel mechanisms that seek answers to “what works” in education and development.

Finally, the favoured metric of cost effectiveness in the report is “how many years of high quality schooling can be bought for an additional $100 spent on each intervention”. No educator believes that learning is a commodity that can be bought in a single transaction at a fixed price. In every real system most of the 100 dollars spent on interventions that are integrated and sustained, will be spent on teachers’ and other staff salaries (and agency costs). Salaries are always most of the recurrent costs of systems. And most of the non-salary costs will be recurrent if interventions are sustained with learning materials and pedagogic technologies that need maintenance and replacement. Recurrent costs cannot be “bought” but they can be financed in sustainable ways (Lewin 2020)[[3]](#footnote-3). Learning is often not linear at least according to Piaget, Vygotsky and Bruner, so “a gain that would take a student in a high quality system twice as long to achieve is one with twice the educational value” is also not a convincing analytic tool in the real world. All gains are not equal.

**Interrogating the Evidence Base**

The report lists about 90 individual evaluation papers and a much smaller set of synthetic studies in an annex and these are used to reach conclusions. About two thirds of the individual studies cited in the report identify the countries where the studies were undertaken. More than half these citations were from only five countries (India (15%), Kenya (14%), Indonesia (7%), Jamaica (7%) and Brazil (7%)). About 25% of the studies were the only studies cited from a particular country. About 24% of studies cited were in South America and 25% in South Asia. As many as 34% were in Sub Saharan Africa countries but 40% of these citations were from one country - Kenya. Strikingly nearly 60% of attributable studies are from LMICs and a further 27% are from UMICs most of which are in South America. Only 14% of attributable studies are from LICs so low income countries appear under-represented.

Several issues stand out.

* The average costs per child at primary level in LMICs are four times those of LICs. In UMICs they are eight times greater. This will make cost effectiveness comparisons in dollar terms look very different for different clusters of countries at different levels of national income. If interventions are externally financed what a dollar will buy will be very different between LICs, LMICs and UMICs. But the measured effect in terms of learning gain may be the same so cost effectiveness will vary widely by country for the sane intervention.
* Studies with large samples were “weighted” to count for more in the ranking of buys. Weighting privileges locations where it is cheap to do large scale data collection e.g. South Asia rather than SSA. Large sample size does not always result in greater reliability, may compound errors in the data, and increases uncertainties in construct validity. Why not give more weight to results of studies that explain more variation in learning.
* Lack of information on how many of the studies cited have longitudinal data or cross sectional time series data that establishes whether performance gains are sustained for more than one cohort of students
* No data on how many of the interventions cited are externally financed, how long ago they took place, and how many are still running after external support has ceased
* Uneven reporting of the variance in achievement that is explained by the interventions and the varying nature of dependent and independent variables measured
* Unclear indication of how many studies have reliable cost data, whether this estimates all direct and indirect costs including donor and government contributions, and how the number of beneficiaries are enumerated.
* Little discussion of the conditions for external validity if results are to be transferred across languages, cultures and national systems and few indications of whether studies have been independently verified
* Covid-19 will have extensive impacts on education budgets and will reshape aspects of effectiveness changing priorities based on pre covid-19 evaluations.

**Destination and Journeys?**

The GEEAP report is built on high quality evaluation reports on interventions that are methodologically homogenous with a strong preference for Randomised Control Trials (RCTs). These do illuminate facts about the effects of experimental programmes in the context of the particular social and cultural systems they were conducted. However, as far as it is possible to tell, none of the studies are replications designed to verify the findings of particular studies in the same context, though some seek to do this across different contexts. RCTs can be good at identifying effects of inputs on outputs under controlled conditions in static rather than interactive conditions. They can give causal insight into closed systems but are not as helpful in understanding open systems like Ministries of Education and development agencies. These are better explored with other research methods.

The GEEAP is planning to commission a series of reviews over the next three years. Those already identified include reviews of school leadership and of mother tongue instruction although these are both topics with an extensive literature. They should start by identifying existing reviews. More particularly there are prominent gaps which the first review leaves largely untouched. Currently there is little of nothing on:

* The impact of Covid which will adversely have an impact of all education systems and may change the cost effectiveness of different recommended buys. It will certainly change the costs and may rewrite the map of priorities.
* Educational financing: This is critical. Sustainable educational development depends on the development of fiscal states that have the ability and political will to self fund mass education systems that provide equitable access to learning.
* Privatisation of provision is barely mentioned though it is widely being considered a policy option. This topic is surrounded by powerful interest groups with an interest in changing the balance of spending
* Surprisingly given the unprecedented levels of investment in girls education this review suggests such programmes apparently have a weak evidence base.
* The current report does not offer a dynamic view of educational problems or “learning”. The assumption that LAYS makes that learning gains are linear and that “a gain that would take a student in a high quality system twice as long to achieve is one with twice the educational value”
* Higher education tends to be disproportionately expensive, is growing rapidly, and is a focus for social mobility into modern sector labour markets is not a subject of discussion. Nor is TVET though this is a perennial topic of policy concern and proposed reform.
* Interventions in high stake examining can be very cost effective in changing pedagogy and are intimately linked to successful curriculum innovation but are not considered any kind of buy, good or bad
* Though there are endemic shortages related to the quality and quantity of education in STEAM subjects but there are no Smart buys identified that relate to investment key curriculum areas.
* The various “Smart buys” are not differentiated by whether they are smart buys for donors or countries. Are donors and countries buying the same things and who are they buying them from? Many of the studies deploy NGO service providers. Who are the principals and who are the agents for the buying and selling?

The epistemology of improving learning in and out of schools is central to the education and development enterprise. Best and other kinds of buys cannot be separated from context because they are embedded in historic experience, normative belief systems, and the realities of resource endowments. What is a Good buy here may be unnecessary or ineffective there. A good pair of shoes is not a good pair of shoes unless they fit. And as astronomers say your best telescope is not the most expensive or the largest, it is the one you use most often. Improving learning may be best approached by asking teachers and students what works for them, rather than imagining top down cookie cutter solutions that have proved elusive for the last 50 years. It may happen in a more resilient way through a myriad of local level investments in professional development and action research based professional development. And it may have most impact on development if interventions are designed with knowledge of life cycle progress though education systems, the zones of exclusion they encounter, and the opportunities that exist to move beyond pricing separate interventions to approaching educational improvement as an open rather than closed system problem.

Interventions must also interface with the global architecture of aid This is a recurrent subject of discussion not least amongst this report’s audience of ”technical staff in Ministries of Education., donor agencies, Local Education Groups and non-profit organisations” One aspect of the architecture illustrates some of the problems. LogFrames entered the development planning world in the 1970s courtesy of USAID. This approach requires specification of Activities, Outputs, Purposes and Goals. USAID. More recently Theories of Change (ToC) have become mandatory since around 2010. These are overlaid on top of requirements for LogFrames and do not substitute for them. These ToC are not really based on theories with propositions about how and why changes take place. More accurately they can be described as plans for change that assume causal pathways that lie within the control and timescale of projects that are undertaken by agents who share the goals identified in LogFrames. The Smart Buys approach steps outside this convention to suggest that there are generic Best buys that are not already embedded in these planning approaches.

The GEEAP has yet to place its theories of change in the public domain. Without this the list of Best buys is denied insight into how the ideas behind such interventions can be converted into durable reforms that are embedded, owned and financed in national systems rather than temporarily grafted on to them. Thus whatever the Theory of Change is that underpins the advocacy of best buys in GEEAP it needs elaboration. Almost all development projects (and rigorous systematic literature reviews), are now required to make their theory explicit so that is a priority for GEEAP. It just might be better to develop a Theory of Innovation (change with a purpose) if the learning crisis is really more of a persistent problem of financing and implementation than an inability to discover “what works”. Learning is a journey more than a destination.

**In Conclusion**

Improving learning may be best approached by inviting teachers and students to identify what works for them, building on what they currently do successfully in the contexts they actually work in. This may have more traction that seeking the top down generic cookie cutter solutions devoid of context that have proved elusive for the last 50 years. This could happen through a myriad of local, district and regional level investments in professional development and action research. Every country has poor children who do learn to read and write. The tragedy is that the opportunities they experience are not universally available to all. What needs to be known is largely known but not necessarily translated into achievement and attainment through the actions of those who need to know. The poorest countries have very different needs to those well on the way to becoming fiscal states beyond the need for external financing. The test of which interventions really are value for money and which are sustainable will be revealed when it becomes clear which ones are adopted without the short term incentive of concessional financing, and which ones persist beyond the lifetime of external project funding.

Disaggregated and community based approaches may have most impact on development if interventions are designed with knowledge of the flow of cohorts of learners though education systems, the zones of exclusion they encounter (CREATE 2012)[[4]](#footnote-4), and the opportunities that exist to manage innovation and learning in open systems, rather than focusing on the price and effects of decontextualised interventions. The SDGs had the ambition of supporting decentralised and endogenously driven development fit for context defined locally within a global framework. This is how they were to be different to the MDGs and how they were to encourage resilience. Durable “better buys” are more likely if locally developed and marketed, purchased from domestic revenue, and designed to be sustainable. Sadly fashions change, sales always finish, and bargains generally have a sting in the tail. In the new development decade it should be different in both poorer and richer countries as more LICs and LMICs become fiscal states able to finance their own educational development. How this can be achieved is elaborated in <https://keithlewin.net/wp-content/uploads/2020/10/Beyond-Business-as-Usual-Aid-and-Financing-in-SSA-Keith-Lewin-x.pdf>

1. This post is a fuller and earlier version of the blog that appeared on the UKFIET website on November 16th, 2020 at <https://www.ukfiet.org/2020/smart-buys-great-sales-and-special-offers-cost-effective-approaches-to-improve-global-learning/>

 [↑](#footnote-ref-1)
2. *Kahneman D, 2012, Thinking Fast and thinking slow*. Macmillan.  [↑](#footnote-ref-2)
3. Lewin K M 2020  **Beyond Business as Usual International Journal of Educational Development**[**https://www.sciencedirect.com/science/article/pii/S0738059320304065?via%3Dihub**](https://www.sciencedirect.com/science/article/pii/S0738059320304065?via%3Dihub) [↑](#footnote-ref-3)
4. Consortium for Research on Educational Access, Equity and Transitions funded by DFID <http://www.create-rpc.org/pdf_documents/Making-Rights-Realities-Keith-Lewin-September-2011.pdf>

 [↑](#footnote-ref-4)