Evaluating sector support using secondary data

Antonie de Kemp
Policy and Operations Evaluation Department
The Netherlands ministry of foreign Affairs[1]

1. Introduction: the quest for the holy grail

Like several other countries, The Netherlands is increasingly moving from project aid to sector and general budget support. This change in policy raises new challenges for the evaluation of the effectiveness of aid, notably at a moment when the measurement of the effectiveness and impact of aid has become an issue in the international debate.

Several papers for the DAC workshop on impact evaluation give an excellent overview of the problems with the measurement of effectiveness and impact. The key issue is the problem of the *counterfactual*: what would have happened in the absence of the intervention(s)? Most (related) issues in the debate arise from this problem:

1. the *attribution problem*: which effects can be attributed to the intervention(s) and in what degree are effects the result of other factors?
2. *selection* effects: intervention group and control group may have different characteristics;
3. the problem of *unobservables*: unobserved variables that are correlated with the intervention(s) and with the measured effects. In that case, the measurement of the effects of the interventions may be biased.

Theoretically, the best way to overcome these problems is randomisation: in a completely randomised design, there is no reason to expect statistically significant differences between intervention group and control group. However, one may be cautious here as well. For instance, in drug testing it is common to make use of a placebo. In other words: the estimate of the effects of the drug may be biased when one doesn't control for the placebo-effect. There are objections to a randomised design as well. First of all, it may not be feasible in practice, because often the projects and programs have already been implemented, without taking into account future evaluation. Secondly, evaluation is a function of a (policy) process and not the other way round. The demands of a randomised design may conflict with the policy or the logic of the intervention.

2. Evaluating sector support

The measurement of the effectiveness of aid becomes more complicated with Sector Wide Approach and General Budget Support (Elbers and Gunning, 2006). Rigorous methods for evaluating impact are designed for projects rather than sector aid or general budget support and at the higher level of aggregation, the interventions to be

[1] This paper owes much to lengthy discussions with Jan Willem Gunning and Chris Elbers (Free University, Amsterdam). I am grateful to comments on a preliminary version from both researchers as well from Henri Jorritsma and Emina van den Berg (IOB).

evaluated are heterogeneous.[2] Multivariate techniques may be used to deal with this heterogeneity, but these techniques require larger samples.

With these challenges in mind, the Dutch Policy and Operations Evaluation Department (IOB) has started a series of impact evaluations. At the moment, the IOB is working on evaluations in two sectors:
- on water and sanitary facilities in selected countries, with a pilot in Tanzania;
- on primary education in Zambia and Uganda.
The pilot of the first evaluation focuses upon a long term program in Tanzania (in the Shinyanga region). The evaluations in Zambia and Uganda are country-wide. They try to assess the effectiveness of interventions in primary education.
The main idea of these evaluations is to assess the impact of interventions using secondary data.

The Tanzania and Zambia studies are carried out by Jan Willem Gunning and Chris Elbers of the Free University in Amsterdam. In both cases, the relevant ministries in the partner countries are heavily involved, as capacity building is an explicit objective of the studies. IOB staff carries out the Uganda study, in close cooperation with the planning department of the Ministry of Education and Sports (MoES) (in Uganda) and the Uganda National Examinations Board (UNEB) and with the technical support of Gunning and Elbers.[3]

3. The Primary Education Study in Uganda

The Zambia and Uganda studies have many similarities, although there are (methodological) differences as well. The Uganda study tries to combine two evaluation functions: the accountability and the learning function. Central questions are:

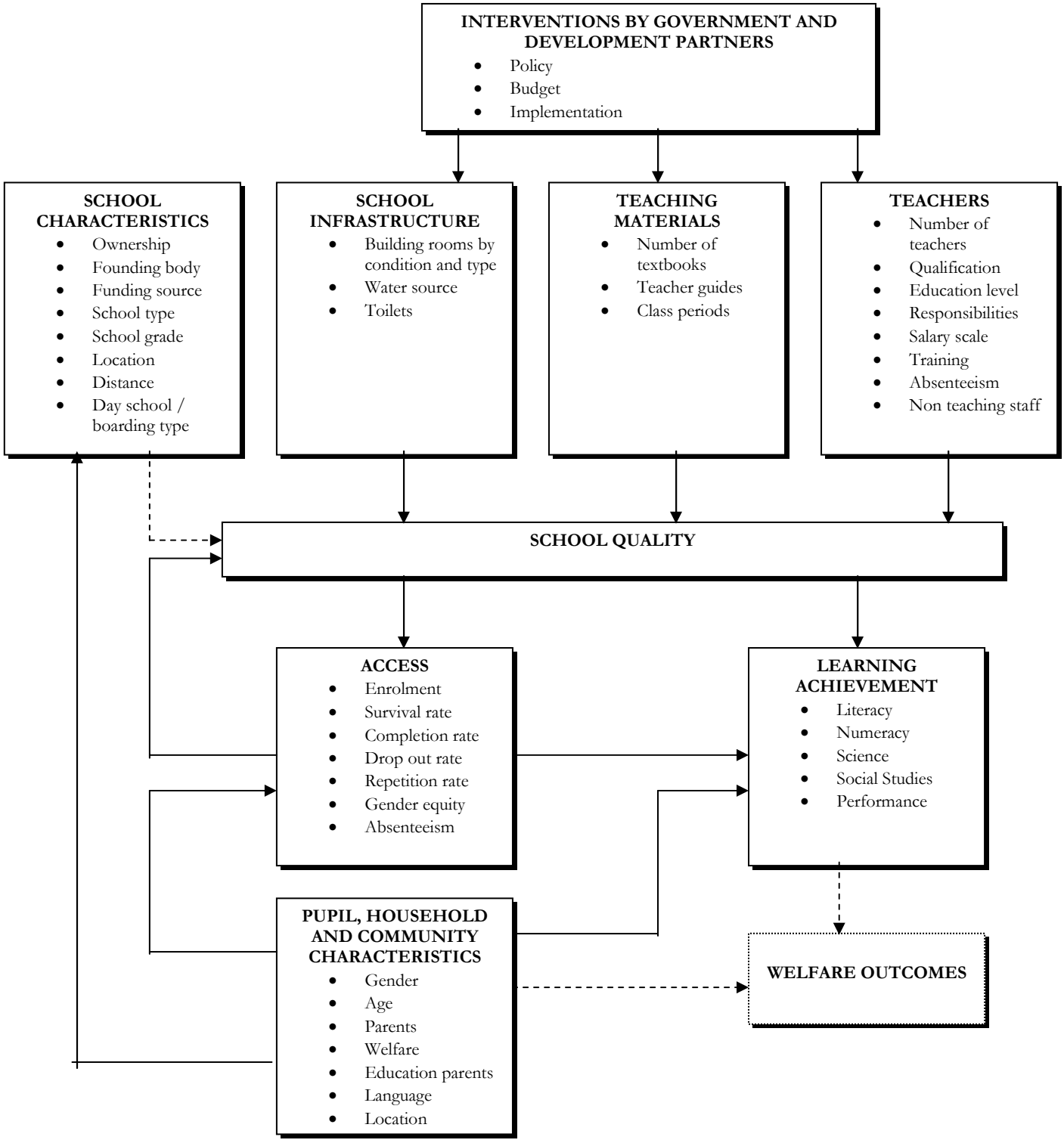
1. In what way have school attendance and learning achievement developed since 2000? The evaluation focuses on:
   (i) Access (enrolment and attendance)
   (ii) Equity / gender
   (iii) Quality (learning achievement).
2. What were the main determinants of these developments?
3. Which interventions have the largest and most (cost-) effective impact on educational outcomes?

IOB supports the World Bank view of the desirability of a theory-based approach (see White, 2006, p. 7-8), even though such an approach introduces new estimation problems (Elbers and Gunning, 2006). The Uganda study uses regression analysis, based on the theoretical model that is sketched on the next page. In the model, you may find input (interventions), output (at the school level), outcome (access and learning achievement) and impact. The impact evaluation focuses at the effects of

[2] The heterogeneity in treatment is a precondition for an impact evaluation at the sectoral level.
[3] Their advisory role does not mean that both researchers may be held responsible for flaws in the research design and analysis in the Uganda study.

interventions on learning and learning achievement (and not on final welfare outcomes).[4]

Figure 1: *Model for impact analysis*



[4] There is a long time span before the welfare effects of primary education become visible. Since the study is interested in the effects of recent interventions, this precludes an analysis of welfare effects. However, the study will try to assess the welfare effects on the basis of the literature. For instance, a change in education levels could be translated to a change in life-time earnings on the basis of a relationship between education and wage levels.

For both (the Zambia and Uganda) studies, the main sources are:
1. school census data
2. test and examination results
3. population census data
4. DHS Education surveys.

In both cases, additional surveys at the districts and schools were needed to get additional financial information, more information on interventions (like the building of classrooms or the acquisition of books) and information on pupil and teacher attendance. On teacher attendance, however, the results are disappointing.

The heart of the analysis consists of linking EMIS data (school census data in the Education Management Information System) to test and examination results at the school level. The (Uganda) study uses data for the years 2000-2005. However, the National Assessment of Progress in Education (NAPE) tests are not based on a panel and only a few schools are sampled more than once. On the other hand, the correlation between test and examination results (the primary leaving exam at the end of the last grade of primary education) appears to be very high. Examination data of about 7.000-8.000 schools have been matched with EMIS data (that is more than 50% for each year). The three databases have been completed with an additional survey of 378 schools in 25 districts. These schools have been sampled from the schools that were in the NAPE surveys in the years 2003-2005.

Although the analysis is at the school level, it is important to take into account household and community (or regional) characteristics. There are first of all large differences between the districts in Uganda (even apart from the rural – urban distinction). Several Northern and Eastern districts are affected by the internal conflict. Many people live in Internally Displaced Peoples' Camps. In the Karamoja region in the Eastern part of Uganda, pastoralists live in semi arid plains. Livestock is the major source of income and the Karimojong see education as a too long investment in comparison with their traditional education. In the Kalangala islands in the Lake Victoria, the fishing communities have a nomadic way of life as well.

The impact evaluation uses two sources to take into account the differences between households and regions. The first one is the DHS Education Survey of 2001. With this survey, it is possible to relate pupil and household characteristics to school attendance. Another source is (a sample of) the Population and Housing census of 2002. Census information is important for the socio-economic differences between regions. The study tries to link the DHS with the other data at the county level. The census data may be linked at the sub-county level.[5]

An important omitted variable in the model is school management. Class size may be endogenous and may be correlated with school management (see for instance Glewwe and Kremer, 2005). However we were able to get information on the quality of management in two districts (about 125 schools).

[5] Uganda has an administrative structure of 77 districts and 146 counties. Each county is divided in several sub-counties.

4. Methodology

Basically, the study uses a regression based approach, with the interventions or outputs at the school level as regressors and access and learning achievement as the dependent variables. There is a huge literature on education production functions and on the effects of educational resources upon learning achievement, which can be summarised with two words: *selection bias* (see for instance Glewwe and Kremer, 2005 and Webbink, 2005).[6] Selection effects create differences between the intervention group(s) and control group(s) and therefore the estimates of the intervention(s) may be biased (White, Sinha and Flannagan, 2006, pp. 3-4). As long as selection is based on observable characteristics, these characteristics may be included in the analysis (see page four for household and regional characteristics). However, often they are not observed.

The Uganda study tries to deal with the issue of unobservables in the following ways:

1) an analysis of the random allocation of interventions on the basis of our own survey and EMIS data;
2) double differencing;
3) the exploitation of natural restrictions;
4) triangulation.

1) *Check for random allocation*. When interventions (for instance the delivery of books or the building of new classrooms) are not correlated with the outcome variables, we assume that there is no reason to expect an endogeneity problem, i.e. that an unobserved variable correlates with the intervention and with the outcome variable. Of course, we are precisely interested in a correlation between the interventions and the outcome variables, but when we include a time lag and compare the interventions in year t with the outcome variables in t-1, these variables do not need to be correlated.
2) In case an intervention variable appears to be endogenous, a way to get rid of the problem is a *double difference* estimate. Of course, this presupposes that the unobserved variable(s) is (are) time-invariant. However, it seems safe to assume that school management and household characteristics do not change very much within one or two years.
3) The *exploitation of natural restrictions*. It is often assumed that the choice of motivated (and probably well educated) parents is correlated with (for instance) class size, as they may tend to send their children to schools with low pupil/teacher ratios.[7] If that case, may estimate the effects of class size for situations where parents do not have a choice (that is in remote rural areas).[8]

[6] Many studies try to measure the impact of class size on learning achievement and many of them come to the conclusion that there is no significant – or even a positive (!) – relationship between class size and learning achievement. The neglect of the endogeneity of class size seems to be an important explanation for these results. However, there is more. Many of these studies have been carried out in western industrialized countries. They do not appear to be very useful for education policy in developing countries. The concept of class size – with a number of children in room with a teacher, stone walls and a roof may be totally different from this concept in rural areas in Sub-Saharan Africa.

[7] It may be difficult to appreciate the consistency of the argument. Well educated parents send their children to schools with low pupil teacher ratios and therefore these schools have better results than other schools and not because the effect of the class-size. But these well educated parents read the literature as well, so: why bother?

[8] This approach is not new and has been applied by Case and Deaton (1998) and Hoxby (2000). See Webbink (2005).

4) *Triangulation*. The study uses different databases and tries to combine data whenever possible and meaningful. Although it is not always possible to link data, we use all the data we've got to get a good grip on the (assumed) relations. For instance, with the DHS Education data, it is possible to analyse the relation between household and pupil characteristics. With the management data (for two districts) it is possible to analyse the effects of (good) management and the correlation of management with the intervention variables.[9] The NAPE test data for 2003 have information on pupil and school (management) and teacher characteristics. Unfortunately, these three surveys are based upon cross section data

There's another way to proceed as well. Schools in one of the districts with management information have been part of a pilot for five years. The intervention has focused its activities mainly on enhancing the quality of education management at both school and district level. A comparison of the results in this district with other districts shows that the pilot appears to be highly successful. A comparison of the schools in this district with comparable schools, using the method of *propensity score matching*, may show more convincingly the success of this pilot.

5. Challenges

Advantages of secondary data are efficiency and (in this particular case) the availability of baseline information. There are disadvantages as well. An important disadvantage, when using secondary data, is that the data are collected for another purpose. So, you will hardly ever get all the data needed. IOB has tried to solve this problem with an additional survey.

The main issue appears to be the quality /reliability of data. For instance, in Uganda it is not easy to give reliable estimates of enrolment rates. According to official figures (derived from EMIS data), net enrolment rates may be much higher than 100%. A part of the large increase in enrolments may (indeed) be the result of the inflation of enrolment figures. An additional survey has been set up to check the reliability of the EMIS data. The linking of data may make it possible to check some data (for instance combining examination data with EMIS data). The enrolment in grade 7 should be consistent with the number of pupils who have registered for the primary leaving exam. In fact, we've found a correlation between (too) high enrolment rates and the ratio between grade 7 pupils and the number of pupil registered for the primary leaving exam. A correlation of very high enrolment rates and low completion rates points to the inflating of enrolment figures as well. The linking of data gives valuable information: in data from UNEB (examination data and test results), children are older than in the EMIS data. This may explain a part of the (too) high enrolment rates. So, you need to check your data over and over again. Therefore, the need, to combine different data sources, seems to be more a blessing than a burden.

As long as there are no systematic error in the measurement of the variables, errors in the measurement are not really a (large) problem. And even a systematic error

[9] It would have been very helpful if this analysis would help to select one or more variables in the EMIS database that could be used as a proxy for the omitted variables. However, that doesn't seem to be possible.

doesn't necessary lead to biased the results (when the error is not correlated with one of the variables in the model). However, when the error term is correlated with one of the variables in the model, this creates actually an endogeneity problem. In other words: a systematic measurement error may be treated as an unobservable.

6. Concluding remarks: better a good estimate than complete ignorance

In one of the excellent papers on impact evaluation, the IEG states "better no numbers than silly numbers". That is right. On the other hand, we must take care not to take the opposite position as well. Partner countries and donors have many questions about the development of education and they need answers. For instance, in Uganda the low (and decreasing) completion rates are a huge problem. We cannot afford to react with the suggestion that we should set up a randomised design with an intervention group and with a control group and that we will give the answer within three or four years. And we cannot urge the ministries to collect management information on a yearly basis without analysing the results and then simply tell them that this information is useless for analysis purposes. Secondary data increasingly become valuable sources for impact evaluations. The challenge is to improve the methods for evaluating impact using these data. Denying their value and insisting on a laboratory approach seems to be a regression.

Literature

- Chris Elbers and Jan Willem Gunning, *Assessing Budget Support with Statistical Impact Evaluation*, Free University, Amsterdam, 2006.
- Paul Glewwe and Michael Kremer, *Schools, Teachers and Education Outcomes in Developing Countries*, CID Working Paper no. 122, Cambridge, 2005.
- Operations Evaluation Department, *Books, buildings, and learning outcomes: an impact evaluation of World Bank Support to basic education in Ghana*, Washington DC, The World Bank, 2004.
- Dinand Webbink, Causal effects in /Education, in: *Journal of Economic Surveys*, vol. 19, no. 4, pp. 535-560.
- Howard White, *Impact Evaluation – The Experience of the Independent Evaluation Group of the World Bank*, Washington, 2006.
- Howard white, Shampa Sinha and Ann Flannagan, *A review of the state of impact evaluation*, Washington, 2006.