

# EDUCOEF

## An Expert System providing District-Specific Diagnostics for Educational Enrollment Problems in Africa

**Jeroen Smits**

Radboud University Nijmegen

[j.smits@fm.ru.nl](mailto:j.smits@fm.ru.nl)

**Version 1, March 20<sup>th</sup> 2011**

### **Introduction**

In many parts of Africa, educational participation of children and especially of girls is too low. Policy makers who want to improve this situation often lack the knowledge needed for developing tailor-made policy interventions. The available scientific knowledge on factors that influence participation is often too crude and too general to be of use in specific situations. Qualitative information on the causes of non-enrollment provided by field workers is often not reliable enough to base policy measures upon.

EDUCOEF is a newly developed Expert System that derives district-level information from available household-level datasets and translates it into a form that is understandable, accessible and usable by policy makers. The Expert System provides them with district-level diagnostics that provides insight into the causes of educational non-participation of children. EDUCOEF can be accessed through the internet and is equipped with an interface through which the information is made available in an understandable form.

In the next sections I will first explain how EDUCOEF works and what can be done with it. Then I will give an overview of the risk factors included in it. The second part of the paper consists of a user guide that provides interested policy makers and other users with the information they need to derive information from EDUCOEF and on the meaning of the coefficients provided by it.

### **Indicators & Coefficients**

Policy makers aiming to improve unfavorable situations with regard to education, health, employment, or other socio-economic outcomes tend to be focused mostly on the development of reliable indicators. Indicators are important because they show how things are and develop over time in a specific policy area. If indicators show that educational participation is relatively low in some regions of a country and that it is higher in other regions, policy makers can use this information to direct their efforts and resources towards the districts where they are mostly needed. Moreover, by monitoring the values of the

indicators over time, they can see how successful their policies have been and, if necessary, change them into a more effective direction.

However, having good indicators is only a first step towards improving unfavorable situations. Indicators learn us *how* things are and *how* they develop over time, but they don't learn us *why* things are as they are and *why* they develop in a certain way over time. For developing effective policy measures, also answers to these why questions are required. Hence, besides reliable indicators we also need insight into the underlying processes. We need a *diagnosis*.

To diagnose educational enrollment problems, we have to look at the micro level of households and their nearby environment, for it is at that level that decisions regarding schooling of children are taken. We thus need household level data. And -- as decisions regarding schooling may be influenced by many factors -- we have to analyze that data with techniques that make it possible to study all relevant risk factors simultaneously. Those techniques generally produce coefficients that show the relative importance of different risk factors for explaining the variation in enrollment. EDUCOEF provides such coefficients for 416 sub-national regions within 42 African countries.

### **How does EDUCOEF work?**

The basic idea underlying EDUCOEF is that undesirable outcomes with regard to educational participation of children are *local phenomena*. The prevalence of such outcomes is supposed to vary among geographic areas and to be related to characteristics of those areas and of the households living there. The characteristics of households that may affect educational participation are called *risk factors*. If we want to develop policy measures aimed at improving an undesirable situation in a given area, we have to know what the values of the various risk factors are in that specific area. With that information, interventions can be developed that address the factor(s) that are most important there.

EDUCOEF provides this kind of information. Using data on educational participation and risk factors from a huge database, EDUCOEF creates and delivers understandable and accessible knowledge on the relative importance of risk factors for educational participation in 416 districts within 42 African countries.

To make this possible, first a basic set of risk factors and context factors was determined, building on the outcomes of earlier research (Huisman & Smits, 2009; Smits & Huisman, 2010). The risk factors include socio-economic, demographic, and cultural characteristics at the household level. The context factors include economic, infrastructural, cultural and political factors at the level of sub-national regions and at the national level. These risk factors and context factors are presented in Table 1.

At the heart of EDUCOEF are multilevel logistic regression models (e.g. Hox, 2002) with indicators for educational participation of girls and boys as dependent variables and the risk factors and context factors as independent variables. These regression models include (cross-level) interactions which ensure that for each outcome district-specific estimates for the risk factors are produced. For each model around 500 interactions have been studied, of which about 100 had to be included in the models to address the variations among districts well.

The model is fed with a multilevel dataset derived from the Database Developing World, a huge data infrastructure developed by the author, in which household level datasets from various sources are brought together, harmonized and supplemented with context information at sub-national and national level ([www.databasedevelopingworld.org](http://www.databasedevelopingworld.org)).

For constructing EDUCOEF, information in the DDW derived from the Demographic and Health Surveys ([www.measuredhs.com](http://www.measuredhs.com)), from the UNICEF Multiple Indicator Cluster Surveys ([www.childinfo.org](http://www.childinfo.org)) and IPUMS International ([international.ipums.org](http://international.ipums.org)) was used for 317,000 children aged 8-12 in 42 African countries. These countries are Angola (MICS 2000), Benin (DHS 2006), Burkina Faso (DHS 2003), Burundi (MICS 2005), Cameroon (DHS 2004), Central African Republic (MICS 2000), Chad (DHS 2004), Congo Brazzaville (DHS 2005), Congo DR (DHS 2007), Cote d'Ivoire (DHS 2005), Egypt (DHS 2005), Equatorial Guinea (MICS 2000), Eritrea (DHS 2002), Ethiopia (DHS 2005), Gambia (MICS 2006), Ghana (DHS 2003), Guinea (DHS 2005), Guinea Bissau (MICS 2006), Kenya (DHS 2003), Lesotho (DHS 2004), Liberia (DHS 2007), Madagascar (DHS 2004), Malawi (DHS 2004), Mali (DHS 2006), Morocco (DHS 2003), Mozambique (DHS 2003), Namibia (DHS 2006), Niger (DHS 2006), Nigeria (DHS 2003), Rwanda (DHS 2005), Sao Tome and Principe (MICS 2000), Senegal (DHS 2005), Sierra Leone (MICS 2005), Somalia (MICS 2006), South Africa (IPUMS 2001), Sudan (MICS 2000), Swaziland (DHS 2006), Tanzania (DHS 2004), Togo (MICS 2006), Uganda (DHS 2006), Zambia (DHS 2002), and Zimbabwe (DHS 2006).

The outcome coefficients of the multilevel logistic regression models contain the information needed for determining the values of the risk factors in a specific district. These values are obtained by multiplying the outcome coefficients with the district and national characteristics in the context database. The district-level coefficients of the risk factors contain the policy relevant information for which EDUCOEF was designed. They show the relative importance of each of the risk factors in explaining the variation in educational participation in the district. This information shows policy makers which risk factor(s) at the household level are most strongly related to educational (non)participation.

To make this information easily available to potential users, EDUCOEF is provided with a web-based tool through which the outcomes are visualized. In the current (first) version of EDUCOEF the coefficients are visualized in a fixed way. In future versions it will be possible to visualize them in different ways according to the preferences of the users, and to use them in a simulation model for developing and testing interventions aimed at improving participation.

### **How to use EDUCOEF?**

The web menu offers the possibility to choose a country from a list of 42 African countries. Then a list of sub-national regions within that country is presented of which one has to be selected. The website then presents two graphs that show the coefficients of a basic set of risk factors for educational nonparticipation. The values of the coefficients are region-specific; they show how important the risk factors are in the selected sub-national region within the selected country. The coefficients are presented separately for boys and girls. Besides values of the risk factors, the graph also presents information on primary school enrolment in the region. A bar with this information is shown on the right-hand side of the graph.

EDUCOEF presents two types of coefficients, coefficients of numerical factors (indicated by 'n' at the website) and coefficients of dichotomous factors (indicated by 'd' at the website). The coefficients of numerical factors show the difference in the chances (odds) of being in primary school between children from households that differ by one step on that factor. For example, a coefficient of 1.5 of the factor 'education mother' means that the odds of being in primary school are 1.5 times (or 50 percent) higher for each step

increase in the educational level of the mother. And a coefficient of 1.2 for birth order means that for each step a child is later in birth order the odds of that child being in primary school are 1.2 times (or 20 percent) higher. The numerical factors are coded in comparable units. They are standardized, which means that a one-step increase is equal to an increase of a standard deviation. This means that a larger coefficient (thus a larger bar) of a factor compared to another factor means that the effect of the first factor on educational participation is stronger than that of the second variable.

The coefficients of the dichotomous (or yes-no) factors generally show how much the odds of being in school are higher or lower for children in the mentioned category (mother employed, father missing, etc.) compared with all other children. The two indicators for father's occupation are however an exception. Their coefficients show how much the odds of being in school for children of fathers with these occupations differ from those odds for children of farmers.

The values of the coefficients are a major diagnostic instrument for policy makers who want to improve educational participation. They show the independent effect of each risk factor on educational participation in the district. Independent means controlled for all other risk factors in the model. Hence, if the effect would be causal it would indicate to what extent educational participation would change with a unit increase in the risk factor. As all risk factors are coded in a comparable way, their graphically presented values immediately show which of the factors would have the largest effects on educational participation in that district.

Although not all risk factors can be influenced equally well by policy measures, knowing which factors in a given problem situation are most strongly related to educational participation is a major step forward compared to the current state of affairs, in which for many districts only information on the *how* of the situation is available and the *why* more or less has to be guessed on the basis of fragmentary and often uncontrollable qualitative knowledge. EDUCOEF aims to provide policy makers with solid quantitative information, based on data on over 300,000 children, that helps them with diagnosing problem situations and shows them directions in which solutions may be found.

### **Variables used by EDUCOEF**

EDUCOEF uses variables at three levels of aggregation: household level, sub-national regional level and national level. At the household level, parents or other caretakers take the decisions about sending their children to school and about how many years of education are considered sufficient for each individual child. The outcomes of these decisions depend on characteristics of the child, of the household and of the larger context in which the household lives.

#### *Household level*

The risk factors at the household level included in EDUCOEF are gender, age and birth order of the children and socio-economic and demographic characteristics of the household in which they live. Gender of the child is an important risk factors for educational non-participation in Africa (UNESCO, 2010). The most important reason mentioned for this is the weak position of women in the region, which is generally associated with the dominant patriarchal culture (Colclough et al., 2000). Given these gender differences, it seems likely that the factors influencing educational participation in African countries will often differ

between boys and girls. EDUCOEF therefore provides gender-specific values of the risk factors.

Important socio-economic factors affecting educational participation of children are income/wealth of the parents, parental education, father's occupation, and employment of the mother. There is broad evidence that children of more resourceful parents are more in school. Income/wealth remains important, even if education in public schools is free, due to the costs of books and uniforms and because of the opportunity costs of the children not being able to work for pay or help in the household or at the family farm (Basu, 1999; Admassie, 2003; Huisman & Smits, 2009).

Children are more in school if the father is in salaried employment. Especially if he works in a non-manual occupation, we expect him to be more aware of the importance of education and therefore to invest more in his children's education (Breen & Goldthorpe, 1997). The opportunity costs of going to school are believed to be more important for parents who are farmers, since they are more likely to expect their children to help out when there is much work to be done, like during harvests.

The effects of employment of the mother are not clear-cut. When the mother is forced to work because of poverty, the daughters may have to take over her household tasks and therefore have fewer chances to go to school. On the other hand, employment of the mother might increase her power within the household. According to the resource theory of conjugal power (Blood & Wolfe, 1960; Rodman, 1972) the degree to which partners can influence important household decisions depends on the extent to which they bring in valued resources into the marriage. This indicates that mothers who contribute to the household income have more influence on family decisions than women who are housewives or unpaid family workers. It seems likely that such more independent women may be better able to create the possibility for their children and especially for their daughters to go to school.

With regard to the educational level of the parents, there is ample evidence that children from better educated parents more often go to school and stay in school (UNESCO, 2005; Buchmann & Brakewood, 2000; Colclough et al., 2000; Smits & Gündüz-Hoşgör, 2005, 2006; Ersado, 2005). Parents who have reached a certain educational level may want their children to achieve at least the same level (Breen & Goldthorpe, 1997). For the educational enrollment of girls, the education of the mother might be especially important (Emerson & Portela Souza, 2007). Mothers who have succeeded in completing a certain level of education have experienced the value of education and know that it is within the reach of girls to complete that level. Therefore, we expect them to use the insights power and insights derived from their higher education to make sure that their daughters get educated too.

Demographic factors included in EDUCOEF are age, birth order, family size, absence of parents, and living in an extended family. The effect of age may differ depending on the number of years of schooling children tend to obtain in the region. In regions where children go to school only for a few years, like many sub-Saharan African countries, educational starting age is often (much) higher than the official starting age of primary education (Huisman & Smits, 2009) and we expect a positive age effect. In countries with high dropout rates on the other hand, like some MENA countries (Smits & Huisman, 2010), a negative age effect can be expected.

With regard to birth order, there is some evidence that under difficult circumstances the cost of high fertility may be borne by older siblings, rather than by the parents (Buchmann

& Hannum 2001; Ejrnaes & Portner, 2004; Dayioğlu, Kirdar & Tansel, 2009). The younger children in such families have more opportunities to go to school, because the older children run the household chores, do the farm work, or contribute to the household income. Besides birth order, also the numbers of brothers and sisters may play a role. Family size often is negatively correlated to educational participation, probably because the available resources have to be divided among more children (Pong 1997; Montgomery & Lloyd 1998; Buchmann & Hannum 2001). However this is not the case in all situations (Chernichovski 1985), probably because more children also means that there are more helping hands at home, which may raise the chance that at least some children can go to school. The same may be true for an extended family. On the other hand, when one of the parents is missing the need for help of the children may be extra high and the chances that they go to school are expected to be lower (Webbink, Smits, De Jong, 2010).

The role of cultural factors (i.e. patriarchy) at the household level is addressed in EDUCOEF by including the age difference between parents and the age at which the mothers of the children got their first child. Both factors are known to influence educational participation of children negatively (Huisman & Smits, 2009; Bradbury, 2007). Level of urbanization of the place in which the household lives – an indicator of both traditionalism of the environment and of local infrastructure – is also included.

### *Context factors*

Characteristics of the larger environment in which the children live is represented in EDUCOEF by characteristics of the sub-national and national context. At the sub-national level we distinguish between four and over 40 (average about 10) geographic areas within the 42 countries in EDUCOEF. The subdivisions have different names in the different countries (e.g. province, district, wilaya, governorate, state, county). For convenience, we refer to them all with the term ‘district’. The context factors are used by EDUCOEF to make the coefficients of the risk factors situation-specific. For creating the situation-specific coefficients, we use indicators of the educational infrastructure, the labor market structure, the degree of modernization, demographic characteristics, and cultural factors.

For educational infrastructure at the district level, the pupil-teacher ratios, percentages of female teachers and percentages of boys and girls in school are used. At the national level, the public expenditure on education and the legal guarantee of free education are included. Economic factors at the district level are the labor market structure (indicated by the percentage of adults working in a white collar job) and the level of economic development (measured by the district percentages of household with a TV and of households living in an urban area) At the national level, GDP per capita and the corruption perception index are included.

Demographic characteristics are family size (average number of brothers and sisters of the school-aged children), percentage of population under 20 (youth dependency rate), and percentages of school-aged children whose mother or father is missing from the household, all measured at the district level.

Cultural factors at the district level are the average age difference between parents of school-aged children, the percentage of these children whose mother had her first child under age 18, and the percentage of households with grandparents from father’s side (indicating the tendency of girls to marry into the family of their husbands). At the national level, ethnic-linguistic fractionalization is used. Other factors included at the national level

are the percentages of parliamentary seats held by women and the prevalence of HIV/AIDS in the country.

Because for most of the countries included in EDUCOEF it is very difficult to get indicators at the district level from other sources, all these indicators except the pupil-teacher ratios were aggregated from the household-level datasets. The national indicators and the district-level pupil-teacher ratios were derived from other sources (like World Development Indicators, reports and websites of national statistical agencies).

## References

- Admassie, A. 2003. Child labor and schooling in the context of subsistence rural economy. *International Journal of Economic Development*, 20.
- Basu, K. (1999). Child labor, consequence, and cure, with remarks on international labor standards. *Journal of Economic Literature*, 37, 1083-1119.
- Blood, P. M. & Wolfe, D. M. 1960. *Husbands and wives. The dynamics of married living.* New York: Free Press.
- Bradbury, B. 2007. "Why do children of young mothers have poorer outcomes?" Presentation at the 2007 LSAC research conference.
- Breen, R., & Goldthorpe, J. H. (1997). Explaining Educational Differentials: Towards a Formal Rational Action Theory. *Rationality and Society*, 9 (3), 275-305.
- Buchmann, C., & Brakewood, D. (2000). Labor Structure and School Enrollments in Developing Societies: Thailand and Kenya Compared. *Comparative Education Review*, 44 (2), 175-204.
- Buchmann, C. and Hannum, E. (2001). 'Education and Stratification in Developing Countries: A Review of Theories and Research.' *Annual Review of Sociology*, 27: 77–102.
- Chernichovsky, D. (1985). 'Socio-economic and demographic aspects of school enrollment and attendance in rural Botswana.' *Economic Development and Cultural Change*, 33: 319–32.
- Colclough, C., P. Rose & M. Tembon. 2000. Gender inequalities in primary schooling. The roles of poverty and adverse cultural practice. *International Journal of Educational Development*, 20, 5-27.
- Dayioğlu, M., Kirdar, M. and Tansel, A. 2009. "Impact of Sibship Size, Birth Order and Sex Composition on School Enrolment in Urban Turkey", *Oxford Bulletin of Economics and Statistics* Vol. 71, Issue 3, pp. 399-426.
- Ejrnaes Mette and Portner Claus C. 2004. "Birth Order and the Intra-household Allocation of Time and Education", *Review of Economics and Statistics* 86(4); 1008-1019.
- Emerson, P.M. and Portela Souza, A. (2007). 'Child Labor, School Attendance, and Intra-household Gender Bias in Brazil.' *The World Bank Economic Review*, 21 (2): 301–316.
- Ersado, L. (2005). Child Labor and Schooling Decisions in Urban and Rural Areas: Comparative Evidence from Nepal, Peru and Zimbabwe. *World Development*, 33 (3), 455-480.
- Gündüz-Hosgör, A. & J. Smits (2008). Variation in Labor Market Participation of Married Women in Turkey. *Women's Studies International Forum*, 31, 104-117.
- Hox, J. (2002). *Multilevel Analysis: Techniques and Applications.* New York, Erlbaum.

- Huisman, J and Smits, J. 2009. "Effects of household- and district-level factors on primary school enrollment in 30 developing countries", *World development* Vol.37, No. 1. pp 179-193.
- Moghadam, Valentine M. (2004). Patriarchy in transition: Women and the changing family in the Middle East. *Journal of Comparative Family Studies*, 35, 137-162.
- Montgomery, M. R., & Lloyd, C. B. 1998. High fertility, unwanted fertility, and children's schooling. In C. H. Bledsoe, J. B. Casterline, J. A. Johnson-Kuhn, and J. G. Haaga (eds.), *Critical Perspectives on Schooling and Fertility in the Developing World*. Washington, DC: National Academy Press, pp. 216–266.
- Pong, S. (1997). 'Sibling size and educational attainment in Peninsular Malaysia: Do politics matter?' *Sociological Perspective*, 40(2): 227–242.
- Rodman, H. 1972 Marital Power and the Theory of Resources in Cultural Context. *Journal of Comparative Family Studies*, 3, pp. 50-57.
- Smits, J. & A. Gündüz-Hosgör. 2005. Effects of family background characteristics on educational participation of girls in Egypt, Morocco, and Turkey. Paper presented at the conference "Urban Children and Youth in the MENA Region: Addressing Priorities in Education", May 16-18, Dubai, U.A.E.
- Smits, J. & A. Gündüz-Hosgör. 2006. Effects of family background characteristics on educational participation in Turkey. *International Journal of Educational Development*, 26, 545-560.
- Smits, J & J. Huisman (2010). Determinants of school enrollment and gender differences in education in 117 districts of six Arab countries. NiCE Working Paper 10-112. Nijmegen center for Economics.
- UNESCO (2005). *Education for All; Literacy for Life*. Paris: United Nations Educational, Scientific and Cultural Organization.
- UNESCO, (2010). *Education For All Global Monitoring Report 2010: Reaching the Marginalized*. Paris: United Nations Educational, Scientific and Cultural Organization.
- Webbink, E., J. Smits & E. De Jong (2010). Hidden child labour: Determinants of housework and family business work of children in 16 developing countries. NiCE Working paper 10-110. Nijmegen Center for Economics.

## **Colophon**

Idea and Development of EDUCOEF: Jeroen Smits  
 Construction and Analyses: Janine Huisman & Jeroen Smits  
 Website construction & design: Wyldebeast & Wunderliebe

## **Acknowledgements**

The development of EDUCOEF was made possible by grant nr. 20100226\_016 of DANS (Data Accessing and Networked Services, [www.dans.knaw.nl](http://www.dans.knaw.nl)). Additional support was obtained from the Nijmegen Center for Economics of the Institute for Management Research of Radboud University Nijmegen and from the WageIndicator Foundation ([www.wageindicator.org](http://www.wageindicator.org)) which sponsored the website. I am grateful to Kea Tijdens for her many useful suggestions.



Table 1. Risk factors and context characteristics used by EDUCOEF

<b>Risk factors at the household level</b>	
Factor	Label
age	Age of child
famis	Father missing from the household
momis	Mother missing from the household
edyrfa	Years of education of father
edyrmo	Years of education of mother
ocfa2	Occupation father lower non-farm (reference "farm")
ocfa3	Occupation father upper non-farm (reference "farm")
wrkmo	Mother employed
wealth	Household wealth (index 10 deciles)
nsist	Number of sisters
nbroth	Number of brothers
bord	Birth order
biolch	Biological child
mchy	Mother got her first child at a young age (under age 18)
nucfam	Nuclear family
pagdif	Age difference between parents

  

<b>Context characteristics</b>	
Factor	Label
whcold	Married men and women aged 30-49 with a white collar (upper nonfarm) job in district
femparn	Percentage of parliamentary seats held by women
fracn	Ethnic fractionalization (national)
corpenn	Corruption perception index (national)
freedn	Legal guarantee of free education (national)
urbd	Urbanisation level of district
tvd	Percentage of households in the district with a tv
gdpcn	GDPpc (national)
under20d	Percentage of population in district under age 20
nsibd	Average number of siblings of children aged 7 to 11 in district
hivn	HIV/AIDS prevalence (national)
edpbd	Percentage of boys in school in district
edpgd	Percentage of girls in school in district
puptrd	District pupil teacher ratio
femtd	Percentage of female teachers in district
pexn	Public expenditure on education (national)
granpard	Percentage of households in the district with grandparents from father's side
agedifd	Average age difference between father and mother for children aged 7 to 11 in district
famid	Percentage of children aged 7 to 11 in district whose father is missing from the household
momisd	Percentage of children aged 7 to 11 in district whose mother is missing from the household
mchyd	Percentage of children aged 7 to 11 in district whose mother had her first child under age 18