Education and country growth models

---

by

Martin Anders Gustafsson

Dissertation presented for the degree of Doctor of Philosophy (Economics) at Stellenbosch University

Supervisor: Prof Servaas van der Berg

April 2014

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Martin Anders Gustafsson

April 2014

Abstract

The over-arching concern of the three parts of the dissertation is how economics can and should influence education policymaking, the emphasis on the economics side being models of country development and the contribution made by human capital.

Part I begins with a review of economic growth theory. How educational performance and country development have been measured is then discussed, with considerable attention going towards conceptual and measurement complexities associated with the latter. An approach is presented for expanding the number of countries whose educational quality can be compared, by expanding the number of linkable testing programmes. This approach, which above all allows for the inclusion of more African and Latin American countries, is one of the key contributions made by the dissertation to the existing body of knowledge. Three existing empirical growth models are examined, including work by Hanushek and Woessman on the relationship between educational quality and income. Part I ends with a discussion on how the economics literature can best be packaged to influence education policymaking. A 'growth simulator' tool in Excel for informing the policy discourse is presented. The production of this tool includes establishing empirically a feasible improvement trajectory for educational quality that policymakers can use and some analysis of how linguistic fractionalisation in a country evolves over time. This tool can be considered a further key output of the dissertation. A basic model for relating educational quality, via income growth, to teacher pay, is presented.

Part II offers an analysis of UNESCO country-level data on enrolment and spending going back to 1970, with a view to establishing historical patterns that can inform education planners, particularly those in developing countries, on how budgets and enrolment expansion should be distributed across the levels of the education system. The analysis presented in Part II represents a novel way of using existing country-level data and can be seen as an important step towards filling a gap experienced by education policymakers, namely the paucity of empirical evidence that can guide decisions around the prioritisation of education levels. Part II moreover arrives at a few empirical findings, including the finding that enrolment and spending patterns have been systematically different in countries with faster economic growth and the finding that historical per student spending at the secondary level appears to play a larger role in development than was previously thought.

Part III contrasts the available economic advice for education policymakers with what policymakers actually appear to believe in. The focus falls, in particular, on four developing countries: South Africa, Brazil, Chile and China. A few areas where economists could explore the data to a greater degree or communicate available findings differently, in the interests of better education policies, are identified. Part III partly serves as a demonstration of how comparisons between education systems can be better oriented towards providing advice to education policymakers on questions relating to efficiency and equity.

Opsomming

Die oorkoepelende fokus van die drie gedeeltes van die verhandeling is hoe die studie van ekonomie beleid in die onderwyssektor kan en moet beïnvloed. Veral belangrik is modelle van die ekonomiese groei van lande en die rol van menslike kapitaal in hierdie modelle.

Die eerste gedeelte van die verhandeling bied ŉ oorsig van die teorie rakende ekonomiese groei. Hoe onderwysprestasie en nasionale ontwikkeling gemeet word, word dan bespreek, met ŉ sterk fokus op die konseptuele en tegniese kompleksiteit van laasgenoemde. ŉ Metode word aangebied waardeur meer lande se onderwysgehalte vergelyk kan word, deur middel van die koppeling van data van ŉ groter aantal toetsprogramme. Hierdie metode, wat veral die insluiting van meer lande uit Afrika en Latyn-Amerika toelaat, is een van die kernbydraes van die verhandeling tot die bestaande korpus van kennis. Drie bestaande empiriese modelle van ekonomiese groei word geanaliseer, insluitende die werk van Hanushek en Woessman oor die verhouding tussen onderwysgehalte en inkomste. Die eerste gedeelte sluit af met ŉ bespreking oor hoe die ekonomiese literatuur optimaal aangebied kan word om beleidmaking in die onderwys te beïnvloed. ŉ Groei-simulasie hulpmiddel in Excel wat die beleidsdiskoers kan vergemaklik word aangebied en verduidelik. Die ontwikkeling van hierdie gereedskap maak dit moontlik om op ŉ empiriese basis ŉ moontlike trajek vir die verbetering van onderwysgehalte te bepaal, wat vir beleidsmakers nuttig kan wees, sowel as ŉ ontleding van hoe linguïstiese verbrokkeling in ŉ land histories kan ontwikkel. Hierdie gereedskap kan as ŉ verdere sleutelproduk van die verhandeling beskou work. ŉ Basiese model van hoe onderwysgehalte en die inkomste van onderwysers deur middel van ekonomiese groei gekoppel is, word ook aangebied.

Die tweede gedeelte van die verhandeling bied ŉ ontleding van UNESCO se nasionale statistieke van lande oor skoolinskrywings en onderwysuitgawes vanaf 1970, met die oog op die identifikasie van belangrike historiese tendense vir onderwysbeplanners, veral in ontwikkelende lande. Die fokus hier is veral op hoe begrotings en inskrywings ideaal oor die verskillende vlakke van die onderwysstelsel versprei behoort te wees. Die ontleding in die tweede gedeelte verteenwoordig ŉ innoverende manier om die bestaande nasionale statistieke te gebruik en kan beskou word as ŉ belangrike stap om ŉ gaping te vul wat deur beleidsmakers in die onderwys ondervind word, naamlik die gebrek aan empiriese gegewens vir besluite oor prioritisering tussen onderwysvlakke. Die tweede gedeelte bied ook verskeie empiriese bevindinge, soos dat die tendense rakende inskrywings en besteding per student sistematies tussen lande met vinniger ekonomiese groei en ander lande verskil, asook dat historiese besteding per student op die sekondêre vlak blykbaar ŉ groter invloed op ontwikkeling het as wat vroeër gedink is.

Die derde gedeelte van die verhandeling vergelyk die advies wat die ekonomiese literatuur aan beleidmakers in die onderwys bied met wat beleidmakers self blykbaar glo. Die fokus val op veral vier ontwikkelende lande: Suid-Afrika, Brasilië, Chili en China. Gebiede word bespreek waar ekonome in die belang van beter onderwysbeleid tot ŉ groter mate data kan analiseer of bevindings op beter maniere kan kommunikeer. Die derde gedeelte kan beskou word as ŉ demonstrasie van hoe vergelykings tussen verskeie onderwysstelsels beter georiënteer kan word om vir die beleidmaker in die onderwys advies te verskaf rakende kwessies van doeltreffendheid en gelykheid.

CONTENTS

The dissertation consists of three parts, listed below. Each part has its own table of contents, figures and tables. A conclusion for the dissertation as a whole and an integrated bibliography appear at the end of Part III.

Education and country growth models

Part I

An evaluation of existing cross-country models and some quantitative applications of relevance to education planning

CONTENTS FOR PART I

Figures and tables

ACRONYMS USED

The following are acronyms used across the three parts of the dissertation. Only acronyms that are used without an adjacent explanation within the text are listed here. For details on the ISO country codes used in many of the tables and graphs, see Table 21 in Part I of the dissertation.

2SLS	Two-stage least squares
AIDS	Acquired immunodeficiency syndrome
BACE	Bayesian averaging of classical estimates
BAU	Business as usual
CGE	Computable general equilibrium
DHS	Demographic and Health Survey
GDP	Gross domestic product
GER	Gross enrolment ratio
GHI	Gross happiness index
GNI	Gross national income
GNP	Gross national product
HDI	Human development index
HIV	Human immunodeficiency virus
IALS	International Adult Literacy Survey
IAM	Integrated assessment model
ICT	Information and communication technology
IDEB	Index of Basic Education Development (from Portuguese *Índice de Desenvolvimento da Educação Básica*)
IEA	International Association for the Evaluation of Educational Achievement
IIASA	International Institute for Applied Systems Analysis
ILO	International Labour Organization
IPUMS	Integrated Public Use Microdata Series
ISO	International Organization for Standardization
IV	Instrumental variable
LFS	Labour Force Survey
MRW	Mankiw, Romer and Weil (1992)
NAEP	National Assessment of Educational Progress
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary least squares
PASEC	Programme for the Analysis of Educational Systems (from French *Le Programme d'analyse des systèmes éducatifs de la CONFEMEN*)
PIAAC	Programme for the International Assessment of Adult Competencies
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
PPP	Purchasing power parity
RSS	Residual sum of squares
SACMEQ	Southern and Eastern Africa Consortium for Monitoring Educational Quality
SDM	Sala-i-Martin, Doppelhofer and Miller (2004)
SERCE	Second Regional Comparative Study (from Spanish *Segundo Estudio Regional Comparativo y Explicativo*)
SNA	System of National Accounts

SSE	Sum of squared errors
SSF	Stiglitz, Sen and Fitoussi (2009)
TIMSS	Trends in International Mathematics and Science Study
UIS	UNESCO Institute for Statistics
UK	United Kingdom
UN	United Nations
UNDP	United Nations Development Programme
UNESCO	United Nations Educational, Scientific and Cultural Organization
US	United States
USD	United States dollar
VAM	Vandenbussche, Aghion and Meghir (2006)
VID	Vienna Institute of Demography
WEF	World Economic Forum
WVS	World Values Survey
ZAR	South African rand

1 INTRODUCTION

Key texts in the economic growth literature identify the economic growth process as a combination of imitation and innovation. Both are necessary, though less developed countries tend to depend more on imitation, as opposed to innovation. This dissertation can be thought of as a combination of imitation and innovation. The sections of the dissertation where imitation predominates take existing analyses and redo them, sometimes using more recent data, they compare the methodologies of existing analyses, and they test the degree to which existing models can assist in specific policymaking questions. The more innovative sections focus above all on using country-level data in new ways that can inform policymakers.

If there is an over-arching question that the three parts of the dissertation deal with it is perhaps the following: *What is the utility of the economic growth literature and models using country-level data for education policymakers and to what extent have these policymakers made use of the available evidence?* Part I of the dissertation is mainly a critical review of existing theory and empirical texts that could be of use to the policymakers. The various sections of Part I are introduced below. Part II addresses what appeared to be a gap in the literature, namely the need for greater empirical guidance with respect to the prioritisation of the various levels of the education system, in particular the secondary and post-school levels. Here, as in Part I, the focus is on models that use country-level data and implement some type of cross-country comparison. The justification for this focus is presented below. Part II is the more innovative part of the dissertation due to its emphasis on new ways of analysing the data.

Part III moves furthest from the over-arching question referred to above, yet the topics it covers are relevant for this question. The focus in Part III is on analysing the education policy directions of developing countries, specifically four fairly diverse developing countries from three continents, within an economic framework of internal and external efficiency. The aim of the analysis is to assess to what degree and how education policymakers have embraced the recommendations of economists, and what this means for the kind of conversation economists should be having with education policymakers. Part III is thus, like the two preceding parts of the dissertation, focussed on cross-country analysis, except in Part III this analysis uses not data points (as in the first two sections), but the institutional architecture and key dimensions of the education systems of the countries in question. To some extent Part III considers if and how policymakers have taken into account growth models in economics, but this is presented within a broader consideration of whether evidence and theory from the field of economics in general is influencing education policymaking. This broader approach seemed more interesting and feasible. The reception amongst policymakers to growth modelling specifically receives attention in Part I of the dissertation.

Turning to the sections of Part I, section 2 addresses the following question: *What are the theoretical roots of economic models that link education to economic growth and how relevant is this theory for education planning?* The focus here is on eight key theoretical models that have influenced the thinking and empirical modelling we see today. The models examined range from Solow's (1956) neo-classical model to recent Schumpeterian creative destruction models. The models selected focus not only on how education and human capital contribute towards economic growth. They also deal with other contributing factors. Yet as will be seen, human capital became an increasingly central concern. Several of these models are from an era when a lack of data to a large degree forced economists to present theoretical (and mathematical) models with no or very little empirical evidence. There is some discussion of the historical developments, in particular the emergence of electronic data and computing, which helped economics adopt a more empirical approach, as well as the arguments that are still made for an economics discipline that is even better at using empirical analysis to test theory, partly so that policymakers can receive the advice they need.

In section 3, the following question is addressed: *How have economists measured human capital and economic growth at the country level and what are some of the key problems and caveats associated with this measurement?* Various attempts, starting in the 1990s, to produce comparable country-level indicators of the years or levels of schooling attained amongst the adult population are evaluated. Moreover, a recent attempt at taking test score data from several testing programmes and normalising these data is discussed. One innovation presented here is a new method for doing this that permits data from a wider variety of programmes to be used. Turning to economic growth, the literature on this is vast and much of the focus is on evaluating an existing and excellent meta-analysis of the literature. To some extent, new data analysis is presented to illustrate a few of the measurement problems. Apart from measurement problems, the welfarist debates around whether economic growth is an appropriate measure of human progress are discussed. Some discussion of the rapidly evolving theory and evidence around economic and social development in a context of climate change (and other environmental pressures) also occurs. These human development issues are of great importance for education policymakers, for a number of reasons, as will be discussed in section 3.

Section 4 offers a detailed analysis of three empirical models, all from the period since 1990, which analyse the relationship between human capital (and other variables) and growth, using country level data. These three models are used as points of departure to assess a number of other similar models found in the literature. Key considerations are: *How robust and reliable are the available models and what do their differing conclusions mean for education policymakers?* The model (or, strictly speaking, set of related models) considered first is that of Mankiw, Romer and Weil (1992). Their augmented Solow model, where human capital is the augmenting factor, represents a milestone within the neo-classical view of country development. Thereafter a model using a so-called Bayesian Averaging of Classical Estimates (BACE) approach proposed by Sala-i-Martin, Doppelhofer and Miller (2004) is discussed. The approach represents an apparently promising, though still not widely used, way of dealing with variable selection problems. Finally, innovative work by Hanushek and Woessman (2009) that challenges the traditional focus on years of schooling in growth modelling and proposes the use of increasingly available data emerging from international testing programmes, is critically assessed.

Section 5 represents the data in ways that go beyond the models of section 4 with a view to providing policymakers with representations that assist in clarifying important education policy matters. A key question is thus: *How can economic models be meaningfully presented to policymakers, and specifically how could this be done with respect to models examining the relationship between economic growth and educational progress?* How the policymaker can use data to underscore the importance of quality improvements in the schooling system needed for economic and social development is examined. Ways in which the magnitudes of certain effects, in particular the effect of qualitative improvements in education on economic growth, can be represented graphically in the interests of better planning receives attention. Given the importance of teachers and their unions in improving the quality of education, how the future quantity and pay of teachers can be explicitly linked to qualitative improvements and economic growth is explored. Here the focus is largely on South Africa, as policy advice must to a large degree be country-specific. Yet the forms of data representation and the policy arguments would be relevant for many countries, in particular developing ones. These forms and arguments are not entirely new and to some extent the section offers a critique of previous attempts to do similar things.

Section 6 provides an overall conclusion for Part I of the dissertation.

The dissertation as a whole draws from a wide range of the literature. This is inevitable if one's ultimate focus is policy advice. Policymakers, at least good ones, are according to Curle (1969) characterised by their multi-disciplinary approach to problems. In terms of fields and disciplines, where does one locate this dissertation? To some extent, the dissertation draws from literature situated outside economics. Despite its name, the ecological economics literature discussed in section 3 falls outside what departments of economics in universities would generally consider their terrain. Part III, in examining education policy positions in various countries, draws from texts in the field of education. However, most of the literature dealt with in the dissertation, roughly 80% of it, is from economics. Within economics, the dissertation draws to a large degree from the literature on the economics of growth and, specifically, cross-country growth modelling. In terms of the micro- against macroeconomic distinction, this places the dissertation on the macro side, though the treatment of macroeconomics is not sufficiently wide or deep to allow one to describe the key focus of the dissertation as macroeconomic. Literature from the field of development economics, in some ways the economics of growth of developing nations specifically, receives some attention, in particular where definitions of growth and development are discussed in section 3. However, it is probably best to locate the dissertation within what has come to be considered an economics field in its own right, namely the economics of education. A possible problem with this categorisation is that the focus within this field, as reflected for instance in the field's pre-eminent journal, the *Economics of Education Review* (started in 1979), is largely microeconomic. Specifically, much of the focus has been on understanding, firstly, the production of educational outcomes within schools, on the internal efficiency side, using institutional data and, secondly, the relationship between education and earnings, on the external efficiency side (viewed from the education perspective), using household data. Yet cross-country analysis has not been absent in this journal[1]. Moreover, economists such as Hanushek and Woessman, whose work inspires much of this dissertation, are considered to be economists of education and Hanushek has published extensively in the *Economics of Education Review*. (Further discussion of the elements that make up the economics of education field is provided in both Part II and Part III of the dissertation.)

What warrants acknowledgement and some justification is the *exclusion* of certain types of economic analysis from the dissertation. The focus of the dissertation with

[1] See for instance Jamison, Jamison and Hanushek (2007).

respect to economic models is on cross-country growth models. Such models seem especially informative for the policymaker wishing to understand the relationship between education and economic growth, or country development more generally. Yet there are other models which appear to throw light on the same policy question. Two such models, or clusters of models, stand out. The first is rates of return to years of education analysis, an area where the work of renowned education economist George Psacharopoulos has played an important role (Psacharopoulos and Patrinos, 2002). Rates of return have been extensively used to guide policymakers. Specifically they have been used by organisations such as the World Bank to provide guidance on what levels of the education system, or what types of education (in particular general versus vocational) should be prioritised for expansion in order to accelerate country development (Psacharopoulos, Tan and Jimenez, 1986). This guidance, which has tended to include the notable recommendation that developing countries ought to prioritise primary education to a larger extent, has been criticised by both policymakers and economists such as Glewwe (1996). Policymakers in developing countries have at times seen the call for a greater prioritisation of primary schooling as an implicit under-valuation of tertiary education and thus a geopolitically controversial call. Glewwe (1996: 283) criticises the World Bank-type policy recommendations based on rates of return in two respects. Rates of return are unable to gauge the positive externalities associated with human capital development, beyond the income returns to individuals. This concern would encompass the policymaker's concerns around innovation within tertiary education research and its link to geopolitical influence. Secondly, Glewwe (1996) uses an analysis of data from Ghana to demonstrate how the poor correlation between years of schooling and actual human capital acquired through the schooling process can render rates of return analyses that use only years of schooling virtually meaningless. He further demonstrates that replacing years of schooling data with data on actual human capital, obtained through test scores, results in a much truer picture of the relationship between education and earnings. In this sense, Glewwe's (1996) contribution with respect to rates of return analyses parallels the contribution of the Hanushek and Woessman analyses (see section 4.4) with respect to cross-country growth models. All overturn, to a notable degree, the credibility of the years of schooling variable in economic analysis. In the case of Glewwe (1996), however, the use of test score data is not able to resolve the matter of unaccounted for externalities, something that cross-country modelling would by its nature account for. This is a key reason why the dissertation focuses on the latter type of modelling. Moreover, a paucity of data in South Africa that would allow one to gauge the relationship between, on the one hand, knowledge and skills demonstrated in tests and, on the other hand, earnings, imposes a serious limitation with respect to rates of return analysis[2].

A second noteworthy model not considered in the dissertation is the computable general equilibrium (CGE) model, which models over time household and firm behaviour and government policy, generally within one country, and their impact on

---

[2] One apparently feasible dataset is the 2008 to 2010 collection of the National Income Dynamics Study (NIDS). This dataset is analysed by Du Rand, Van Broekhuizen and Von Fintel (2011), who find that considerable noise in the test data limit their utility for a rates of return analysis. Branson and Leibbrandt (2013) circumvent the problems with the test data by linking the NIDS data to measures of secondary school examination results and produce an analysis which is arguably the most promising South African analysis to date of the conditional relationship between educational outcomes, on the one hand, and earnings and employment probabilities, on the other.

macroeconomic variables such as growth. A common use of the CGE model is to gauge the impact of changes to tax policies on macroeconomic variables. Infrequently, education variables have been included in this type of model. Hong-Sang Jung and Thorbecke (2001) offer an example of a model with education variables using data from Zambia and Tanzania and with an education policy focus. In their modelling, done separately for the two countries, changes in government's education expenditure are tracked, which in turn are seen to influence the skills level of labour supply. The analysts conclude that increasing education expenditure would lead to faster economic growth. There appears to be little discussion in the literature of the merits of a CGE approach for assessing the impact of education policy changes. What one can safely assume, however, is that not accounting for qualitative educational outcomes and the efficiency of education institutions will bias any CGE model in a manner that would make any education policy conclusions problematic. CGE modelling is a highly specialised field within economics. Exploring credible ways of incorporating government's education decisions within this model would require a dissertation of its own. For this reason CGE models are not considered within this dissertation.

Quite apart from any debate on what areas of economics to draw from, and what quantitative models to use, when attempting to make a contribution to education policymaking, there is a debate, a curious and often infuriating one for many economists and education planners, around whether economics and quantification have a role to play at all in education policymaking. This debate is described, to some extent, in Part III, when certain academic education texts and teacher union policy documents are discussed. The debate is often completely shunned by economists, though Psacharopoulos has entered the fray on at least one occasion. Psacharopoulos's (1990) view of the debate is reflected in the following extracts from his article in the *Comparative Education Review*, a journal he was in fact criticising:

> If you are reading this, you must be associated with the field of comparative education. Are you a "relativist," "hologeistic," a "neopositivist," ... or, perhaps, "morphogenetic?" ... Before you decide, you may wish to ask yourself a question: Does it really matter? I think not ... What matters is your position on a substantive issue. ... I have expressed uneasiness about the contents of the two major comparative education journals, *Comparative Education Review* and *Comparative Education.* My conclusion is that the articles in … these journals are overly descriptive, in the sense that they provide long, nonquantitative accounts of the educational system of a single country. Seldom are the papers analytical, in the sense of statistically testing hypothesized relationships. As a result, few comparative lessons can be drawn to assist decision makers in educational planning.

Importantly, Psacharopoulos was not questioning the utility for education policymaking of the academic field known as comparative education, but rather arguing that it ought to be better focussed. In fact, Psacharopoulos (1989) puts forward an example of what good comparative education analysis, in his view, should look like. Part III of the dissertation attempts to do something similar. Comparative education and the kind of cross-country modelling seen in Parts I and II ought to be complementary areas of study. The latter covers a sufficient range of countries to allow for meaningful and robust patterns in the data to be detected, whilst the former narrows the focus to a few countries, thus permitting in-depth analysis of policies and the influence of specific national contexts.

Whilst there is sympathy in this dissertation for the frustration expressed by Psacharopoulos in the above extract, there is also an attempt to uncover the institutional and ideological explanations for ‘anti-economist’ views, and also to demonstrate that if the economic approach to understanding education systems is pursued with sufficient care and sufficient recognition of context, misunderstandings can be minimised and quantification can be made appealing.

2 AN OVERVIEW OF KEY THEORETICAL MODELS

2.1 Introduction

Theories of economic growth in the last fifty years have mostly been expressed through mathematical formulas accompanied by a narrative, but different economists tend to use slightly different terms and mathematical notation. To assist the discussion presented below, the mathematics and language used in Aghion and Howitt's (2009) growth textbook is used. This is a textbook written for graduate students and professionals, and is a reworking of the authors' original 1998 textbook (Aghion and Howitt, 1998). The selection, sequencing and assessment of models presented here are also influenced by this book. At the same time, this section is not a summary of Aghion of Howitt (2009). The education policy concerns of the dissertation influence the selection and interpretation of the models, and for certain models the original texts, rather than Aghion and Howitt (2009), were used as the principal source. The key original texts referred to by Aghion and Howitt (2009) were all consulted.

Robert Lucas, author of one of the models discussed below, refers to economic theory as 'an explicit dynamic system, something that can be put on a computer and *run*' (original italics) and 'the construction of a mechanical, artificial world, populated by the interacting robots that economics typically studies, that is capable of exhibiting behavior the gross features of which resemble those of the actual world'. In this section, the criterion that a theoretical model should be implementable on a computer is taken seriously. All eight models are simulated in Excel, and results in the form of graphs are discussed. Tiny fabricated datasets and not actual economic data were used, partly because the models studied, being theoretical, are often not replicable with real world data. To illustrate, the discussion that follows will show how notions of duration and lags implied by the models are not realistic. This is not necessarily a problem if one considers one is dealing with an 'artificial world' designed to 'resemble ... the actual world'. To use the terminology of McCloskey (1994: 42), economic models are metaphors. Both the smallness of the fabricated datasets and the fact that simulating the models often required adjusting the data through an optimisation process, made Excel an appropriate tool to explore the models. Moreover, an aim in this section is to present the eight models in as standardised a fashion as possible, to facilitate comparison between the models.

A short look at some pre-twentieth century thinking seems in place. In Adam Smith's (1991) *Wealth of Nations*, originally published in 1776 and often considered the original treatise on classical economics, it is investment, the division of labour and (to some extent) colonial exploitation that bring about a growth of national income. Investment allows for more capital intensive production and the division of labour allows for specialisation, all of which result in more efficient production. Investment in human capital is explicitly recognised (Smith, 1991: 247):

> [Fixed capital includes] the acquired and useful abilities of all the inhabitants or members of the society. The acquisition of such talents, by the maintenance of the acquirer during his education, study or apprenticeship, always costs a real expense, which is a fixed capital and is realised, as it were, in his person.

Marx, though not primarily known for his theory of economic growth, did put one forward, and the nature of his theory influences his better known historical determinism. Like Smith, Marx views investment, the division of labour and colonial exploitation as growth-enhancing. However, human behaviour in the growth process is rather different in Marx in at least two important ways. Firstly, by allowing for high levels of human exploitation in his theory, Marx permits capitalists to run extraordinarily high social risks to the extent that capitalism itself collapses (Mandel, 2002). This part of Marx is of course well known. What is less widely discussed is the fact that in Marx labour, whilst specialised, remains rather undynamic with respect to its level of knowledge and skills, or its human capital. Workers do not invest in their own human capital (or that of their children) to improve their productivity and wages, nor do capitalists improve the human capital of their labour to increase their profits. Even a basic awareness of investment in human capital, as reflected in the above extract from Smith, seems absent in Marx. The absence of rigorous attention to human capital in Marx is arguably a key factor explaining his historical determinism and his inability to foresee the ability of capitalism to adapt and gain acceptance in a sufficiently large portion of society to survive in ways he did not foresee.

A further difference between Smith and Marx is noteworthy, given what growth models have to say about competition between firms. For Marx, market dominance by monopoly firms is inevitable. Smith, who regards monopoly as having a negative effect on welfare, sees no such historical trend. We shall see that recent growth models focussing on imitation and innovation consider monopolisation as a positive trend, and in some senses inevitable, though in a completely different manner to Marx.

The fact that Marx and Smith should have had such vastly different views of the role of human capital in economic development and that both should have exerted such a strong influence on governments over an extended period of time, underlines the importance and power of formally stated theory on human development, even where this theory is not empirically based. Education policymakers are guided by various theories on the contribution made by education towards the development of society. The more economically and sociologically oriented theories that guide the education sector are sometimes studied formally within philosophy of education classes, which are mainly offered by university faculties of education and teacher training institutions. At least three theories commonly found amongst education experts, other than the human capital model economists would be familiar with, are described by Gilead (2012), writing in the *Journal of Philosophy of Education.* One theory is in a sense a non-theory, an inward-looking view of education coupled with a notion that there is no need to dwell on the relationship between education and social development. Another theory explicitly rejects the human capital theory of the economists on ethical grounds, arguing that its portrayal of human subjects as contributors towards economic growth is de-humanising. Theodore Schultz, 1979 Nobel laureate in economics for his work in development economics and one of the earliest proponents of the term ‘human capital’, sees ethical concerns as a reason why economists themselves took so long to incorporate human quantity and quality into economic models:

> The mere thought of investment in human beings is offensive to some among us. Our values and beliefs inhibit us from looking upon human beings as capital goods, except in slavery, and this we abhor. ... to treat human beings as wealth that can be augmented by investment runs counter to deeply held values. (Schultz, 1961: 2)

Thirdly, some educationists reject the determinism they see in the economic notion of human preferences, specifically preferences for maximum earnings. This, the argument goes, denies the role played by education systems in shaping and changing human preferences. Gilead (2012) finds the latter objection to human capital theory the most interesting one, partly because it is most likely to lead to a constructive and mutually educational conversation between educationists and economists. This is a conversation that Gilead (2012: 114) believes is lacking amongst philosophers of education:

> By exploring economic ideas, philosophers of education, for their part, can render their work more relevant for policy-making and hence play a greater role in shaping it

In examining the theories of economic growth in this section, some attention will go towards understanding how these theories might be received by education policymakers and how they could constructively challenge policymakers to question pre-conceived notions of education's role in society.

Theoretical models are dealt with in this section, separately from the often related empirical models discussed in section 4, which comes after section 3 has introduced the data used in the empirical modelling. This separation between the theory and empirical modelling is followed for two reasons. Firstly, this allows one to deal with conceptual issues arising from the theoretical models in a more integrated way and, similarly, data issues arising from the empirical models can be dealt with more integrally. Secondly, the links between theory and empirical analysis in the growth literature are arguably weaker than what one might expect. Many theorists, in particular the ones writing before the 1990s, did not do any empirical testing of their theories. But even when the same economist presents both theory and empirical testing, the structure of the two is sometimes rather different.

The order of the eight models discussed below is only roughly chronological. The oldest model, Solow's neoclassical model, comes first, but this is followed by a fairly recent vindication of Solow's model, by Mankiw, Romer and Weil. Then comes a discussion of the basic AK model, an attempt to move beyond the confines of neoclassical economics, followed by a specific and recent formulation of the AK model, by Lucas. Next is the product variety model, which introduces the role of firms in the growth process. Lastly, three Schumpeterian models, the Nelson and Phelps model, a 'creative destruction' model and a model with two levels of human capital, are discussed.

Some consideration of McCloskey's arguments relating to the relevance of economic theory is informative at this point. In *Knowledge and persuasion in economics,* McCloskey (1994: 127-145, 164-178) argues that the surge in 'mathematical formalism' beginning in the 1950s, whilst initially a healthy response to insufficient rigour in economics, ended up becoming an obsession that obscured the actual purpose of economics. To illustrate the trend, McCloskey points to statistics such as the increase in the percentage of pages in the *American Economic Review* 'containing a mathematical expression without empirical use' from 2% to 44% between 1951 and 1978 (McCloskey, 1994: 127). McCloskey would support some of this trend, and some of its proponents, such as Solow, who saw mathematics as a means to encapsulate, clarify and develop further the largely narrative approach to economic argument that had predominated. But, according to McCloskey, mathematical formalism was also driven by less noble motives, and hence went too far. One such motive was the need to make economics departments at universities seem more worthy of support through making economics more 'scientific', and hence more like rapidly growing scientific departments. McCloskey has examined this notion of 'science' at length and argues that the economists were fundamentally misguided. The science departments at universities, she argues, are in fact more pragmatic and less mathematical than the economics departments. The economists misunderstood 'science'. McCloskey rejects the argument, put forward by some economists, that economics was forced to become mathematical and relatively abstract due to a shortage of experimental data. She points to geology, which is similarly lacking in experimental data, but has remained more faithful to solving practical problems. Another motive is a cultural one. Many economics departments (she is referring almost exclusively to those in the US) are in the grips of an 'extreme aestheticization' in which the elegance of mathematical theory becomes valued more than inelegant dealings with historical, social and political realities (McCloskey, 1994: 167). McCloskey would like to see economics become more practically oriented, focussing on more empirical analysis and simulation using real world data, and within the area of empirical analysis focussing more on the magnitude of effects and less on whether an effect exists at all (these points relating to empirical analysis receive considerable attention in this dissertation). It is important to bear in mind that the arguments mentioned here with respect to mathematical theory in economics are from a book published in 1994, when access to computers, and skills to use them, were still relatively rare amongst departments of economics staff and their students. Computers have undoubtedly shifted economics in a more practical and empirical direction, and one presumes McCloskey would approve of this. Yet her arguments against excessive mathematical formalism are a useful backdrop against which to examine the evolution of the theory of economic growth, and to inform the conclusion to this section.

2.2 Solow's neoclassical model

Solow's model, like several of the theoretical models receiving attention, are based on a Cobb-Douglas form, so a short review of this form is provided. Cobb and Douglas (1928) provide a rare example of an early empirical examination of growth, though their focus is on growth in production, and in one country, the United States. The Cobb-Douglas form can be considered one of two basic, and to some extent competing, forms in the growth literature, the other being the Harrod-Domar form, which is discussed below (Frankel, 1962). The Cobb-Douglas form can be represented as follows (Cobb and Douglas, 1928: 151; Sydsæter and Hammond, 2002: 378):

$$Y = AK^{\alpha}L^{\beta} \tag{1}$$

The exact meanings of the values may differ somewhat, but generally Y is income or output during some period of time which is typically a year, K is the stock of physical capital in the economy, L is the stock of labour (or the flow of labour time) and A is the level of technology or productivity. A which, as will be seen, became a central concern for growth economists, thus partly reflects the nature of K (better technology and productivity is often a feature of the machines people use) and partly the nature of L (technology and productivity enhancements require more able labourers). Typically, the exponents α and β add up to 1 or unity and each is greater than zero. This means that constant returns to scale hold. If AK and L are doubled, then Y doubles in value.

However, if, say, L is kept constant, then Y is a decreasing function of K, meaning in this instance there are diminishing marginal returns to K. Each additional unit of K is associated with a smaller increase in Y until the increase in Y is barely differentiable from zero. We can express this in terms of derivatives:

$$Y'(K) > 0 \text{ and } Y''(K) < 0$$

The first derivative of Y as a function of K is never less than zero because more capital can never reduce income. But the fact that returns are diminishing is captured in the fact that the second derivative of Y as a function of K is always negative. On the other hand, if *both* K and L are increasing at, say, 1% per year, Y will increase by 1% a year, no matter what values the exponents have, as long as they add up to unity, and the second derivative will equal zero:

$$Y''(K) = 0$$

Solow's model is neoclassical in the sense that it implicitly assumes there is perfect competition, profit maximisation by firms and rational behaviour amongst all economic agents. Solow's model is sometimes referred to as the Solow-Swan model due to the coincidental publication in the same year, 1956, of a similar model by the Australian economist Trevor Swan (Solow, Nobel laureate for his work on the modelling of economic growth, is American). Both Solow (1956) and Swan (1956) presented purely theoretical models, with no empirical application. The form used was the Cobb-Douglas form with the sum of the exponents explicitly being equal to unity (Aghion and Howitt, 2009: 23; Solow, 1956: 84-85; Swan, 1956: 335):

$$Y = AK^{\alpha}L^{1-\alpha} \tag{2}$$

K is determined endogenously through the next formula, where s refers to saving, or conceivably what the System of National Accounts (SNA) calls the 'gross savings rate', and δ refers to depreciation, or 'consumption of fixed capital' in the SNA. The value of the addition to the stock of capital from one year to the next is equal to the amount of income from the previous year saved and invested in new capital stock minus the value of the capital stock that has depreciated.

$$\dot{K} = sY - \delta K \tag{3}$$

In many applications of the neoclassical model, and for the discussion that follows, s and δ are assumed to remain constant over time. Swan (1956: 345), whilst not including δK explicitly in his model, does make important points about depreciation. The capital stock, he argues, is not 'instantaneously adaptable' and it is thus not enough to consider normal depreciation in the sense of capital that ceases to function, but also loss of capital due to equipment and machines becoming redundant as the technology of production progresses.

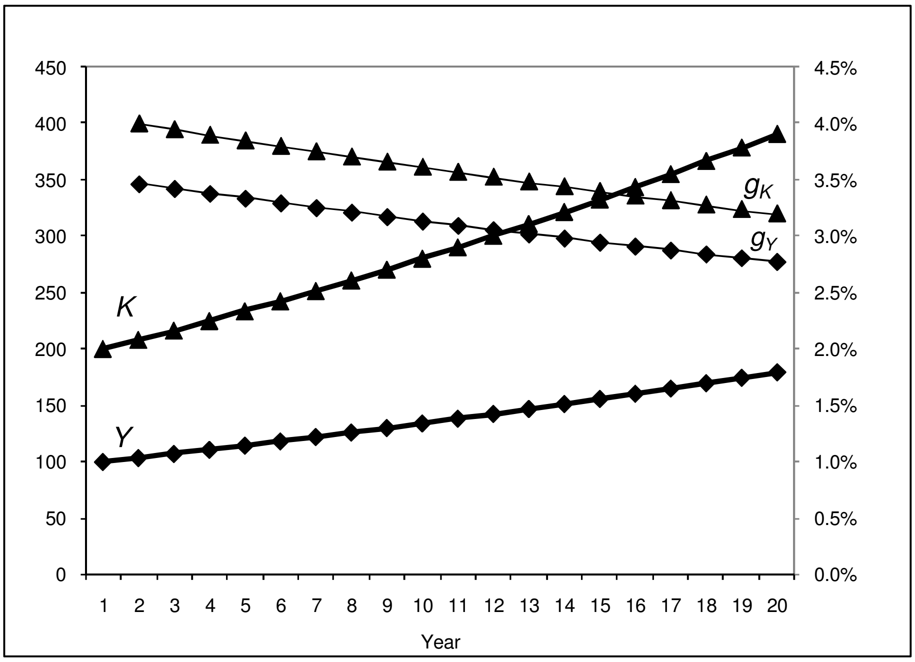
A is the 'productivity' (Aghion and Howitt: 2009: 13) or 'technological change' (Solow, 1956: 85) or efficiency whereby capital and labour are converted to income. As mentioned earlier, A could encompass both features of L, such as level of education and skills, and features as K, such as the quality and efficiency of equipment. The Palgrave Dictionary of Economics (Griliches, 2002) describes productivity as being 'embedded in the current range of equipment and in the training of the existing work-force'. However, Solow's model, unlike many of the later models, implicitly considers A being attached primarily to K.

Simulating Solow's model in Excel requires at least some realism in the numbers if one is to avoid nonsensical results. However, one arbitrary value used for Figure 1 below and in many subsequent graphs is the value 100 as the starting point for income Y. This was done to facilitate comparison across the graphs. For Figure 1, recent South African values[3] were used to inform the starting value of K, relative to Y, so in the starting year K is about double the value of Y, or 200, implying a capital/output ratio of 2.0. The gross savings rate s is 0.2 and consumption of fixed capital δ is 0.06, more or less the values for South Africa. A, which refers to an abstract concept outside the national accounting system, was assigned the value 1. Labour L was also assigned the value 1 for the sake of simplicity (an approach frequently followed in the literature and implicitly followed in many of the models described below where L is omitted). The exponent α attached to K was assigned the value 0.87 so that equation (2) would balance[4]. The following application of equation (3) was followed in the simulation:

$$K_t = K_{t-1} + sY_{t-1} - \delta K_{t-1} \qquad (4)$$

The simulated model over twenty years can be illustrated as follows:

Figure 1: Solow's neoclassical model



The position of the symbols indicate whether the curve should be read against the left-hand or right-hand vertical axis. Thus K and Y should be read against the left-hand axis and g_K and g_Y against the right-hand axis. It is noteworthy that the simulation

[3] From South African Reserve Bank (2009).

[4] As Aghion and Howitt (2009: 58) point out, although in an economy that is in a state of equilibrium the exponent for capital α should be close to the savings rate (around 0.2 for South Africa), its true value is often much higher, between 0.7 and 1.0 (as suggested in the simulation described here), which is indicative of externalities arising out of capital accumulation.

results in growth rates for Y, or g_Y, that are realistic. They begin at 3.5% and drop to 2.8%. The following patterns can be observed:

- In this simulation growth in the capital stock g_K is greater than growth in income g_Y (4.0% against 3.5% initially). This means that over time the capital/output ratio increases, from 2.00 to 2.18 over the 20 years.

- Both g_Y and g_K are falling, and falling at the same rate (the slopes of the g_Y and g_K curves are the same).

Plausible changes to the starting values do not change these patterns substantially, but the following can be noted:

- If the starting values of Y and K are multiplied by a factor, and L is multiplied by the same factor, g_Y and g_K remain the same, implying that the size of the country has no bearing on growth.

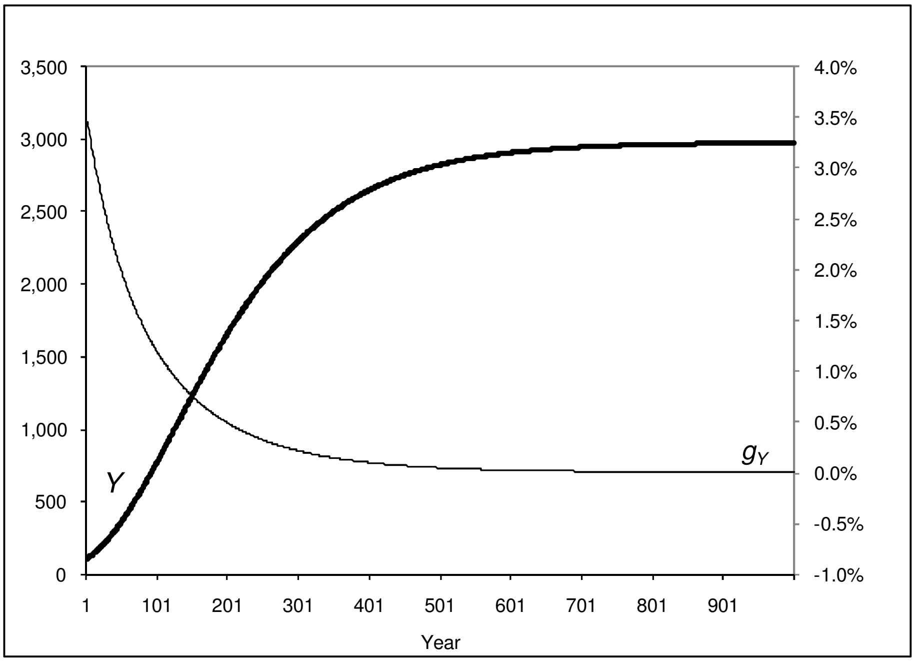
- If the savings rate is reduced by multiplying it by a factor i, where i is less than 1, g_Y and g_K drop by factor that is even less than i. For instance, reducing the savings rate to 0.15 decreases initial growth in Y to 1.3%.

- If the capital/output ratio is reduced, growth in K and Y rise sharply. For instance, reducing initial K to 150, so the capital/output ratio becomes 1.5, results in an initial g_Y of 6.7%. In this adjustment, it is necessary to increase the value of α slightly from the 0.87 referred to above to 0.92. If α is not adjusted in this manner, the model does not balance in the sense that one obtains a strange drop in Y between year 1 and year 2. That less capital should be associated with higher growth seems at first counter-intuitive. However, a country that has a relatively high starting income relative to its capital stock must be very efficient at converting capital into income, hence the higher value for the coefficient α, and the high growth must be understood as the product of high efficiency.

- If the adjustments from the previous two bullets are effected simultaneously, in other words if both the savings rate and the capital/output ratio are reduced, then g_Y and g_K decrease, but only slightly (starting g_Y becomes 3.7%). This means that a country that has less capital stock because it saves less will grow at a lower rate. This makes intuitive sense. In this scenario, α must be adjusted upwards to the extent described in the previous bullet for counter-intuitive jumps between years 1 and 2 to be avoided.

In the long run, what will happen to g_Y? Figure 2 illustrates the simulation from Figure 1 in the long run. Growth in Y decreases to 1.0% in year 118, 0.5% in year 196, and approaches zero but never reaches zero, meaning $Y''(t)$ approaches zero in the long run. The graph suggests that the 'steady state' of Y of around 3,000 is reached after some 600 to 700 years.

Figure 2: Solow's neoclassical model in the long run



It is also possible to calculate the steady state level of *Y* mathematically[5]:

$$Y = \left(\frac{s^{\alpha}}{\delta^{\alpha}} \right)^{1/1-\alpha} \qquad (5)$$

Different starting values for *K* and different savings rates *s* result in different steady states. If our adjustment of the savings rate to 0.15 is applied, *Y* reaches a steady state at around 450 in more or less the year 500. If our adjustment of initial *K* to 150 is applied, the steady state is at around 800,000 by the year 1000 (and arguably the steady state has not been reached yet). The combination of these two adjustments produces a steady state at around 30,000 in the year 1000.

So what does all of this mean for the theoretical country? Above all it means that there is a steady state level of *Y* in the long run, but that this steady state depends on the size of one's initial capital stock, its productivity (as reflected in α), and the extent to which one restocks it. Countries with the same initial values for these three variables must end up at the same steady state of *Y*. Though this implies that countries may not all converge to the same steady state, as not all would have these three starting values, the possibilities for non-convergence are generally considered so restricted that Solow's model is often said to predict convergence amongst all countries to the same or at least a very similar steady state. Importantly, using plausible values results in the steady state being reached very far into the future, often after hundreds of years. Though the fact of a steady state is referred to as a key feature of neoclassical models, the time required for this to occur is seldom made clear[6]. From the perspective of what could happen within one or two generations, perhaps a more important pattern in the Solow model is that there are diminishing returns to

[5] This is derived from a similar formula in Aghion and Howitt (2009: 289) dealing with the Mankiw, Romer and Weil model.
[6] For instance Aghion and Howitt (2009) do not explain this.

capital. Such diminishing returns over time would not be a sign of failed economic policies, but rather of dynamics inherent in the productivity of capital. These dynamics are not difficult to grasp intuitively. If one has no or very little physical capital (be this in the form of ploughs, brickmaking machines or computers), then adding some makes each person more productive. But eventually, the addition to the capital stock used by a finite number of people will not change productivity to any significant degree (what will an economist do with five computers?).

Two provisos must be noted. An increase in labour *L* means that the diminishing returns for *total* additions to the capital stock can be slowed down, but per worker diminishing returns will remain the same. Moreover, according to the model, total income can grow infinitely as long as population and labour grows infinitely. The second and more relevant proviso is that if *A* were to increase infinitely, so would growth, even growth in income per capita.

The Solow model says almost nothing about human capital and, by implication, education, with the possible exception that the model implies education systems should teach the young to save. As mentioned above, one would expect human capital to reside within *A*. Apart from human capital, another element of *A* is not specified in Solow's model (though it is in the Schumpeterian models that will be examined). In Palgrave (Griliches, 2002), productivity, or *A*, is a function of both 'the average set of recipes for doing things' and 'the currently known *best* way of doing things' (own italics). Put differently, productivity is a function of two things: the extent to which we have adopted the best technologies available and what technologies we have not adopted yet, though they are available for adoption in the near future because they have been tested elsewhere. A further element of productivity (also from Palgrave and also not dealt with in Solow's model) deserves mention. Productivity is enhanced by human capital, by the available capital stock, but also by 'the current organizational arrangements for using these resources in production' (Griliches, 2002: 1010), whether these arrangements be at the micro level, in for instance the managerial culture of firms, or at the macro level, in for instance the governance of the country. Clearly, much of the work in growth modelling should be, and indeed is, oriented towards unpacking *A*.

Solow's model has served as an important point of reference and target for critique for subsequent theorists. Apart from the absence of an explicit human capital variable, the element of his model that has probably attracted the most criticism is diminishing returns to capital (and this is of course a fundamental element). Romer (1986: 1008), for instance, in arguing against the neoclassical growth paradigm, indicates that the data do not support convergence towards a steady state of income resulting from diminishing returns to capital. Using historical country-level data going back to 1700 for some countries, Romer demonstrates that contrary to the predictions of the neoclassical model, economic growth tends to increase over time.

The rise and fall of Solow's neoclassical model was part of a larger and highly significant trend in policy thinking. Development economists and institutions such as the World Bank behaved for many years in ways that reflected a strong belief in the power of physical capital investments, and relatively weak attention to the role of human capital. As discussed by Emmerij (2002), in his historical account of international development aid, up until the strong shift towards investments in social programmes, in particular education, in the 1970s, development agencies had assumed that raising investments in physical capital and focussing on strengthening niche industries were the most efficient ways of stimulating development. They were not entirely incorrect, as 'economic miracles' in the 1960s in countries such as Brazil seemed to demonstrate. However, questions about the sustainability of this type of development were raised. In the case of Brazil, rapid industrialisation did not appear to alleviate poverty as expected and income inequalities worsened.

Though the development economics literature began elevating the role of human capital from the 1970s, it was only in 1992 that Mankiw, Romer and Weil re-designed Solow's model in order to incorporate an explicit human capital variable. This delay was partly due the fact that the required data for the empirical part of Mankiw, Romer and Weil (1992) only became available towards the end of the 1980s.

Though in the economic literature a model of industrial development and GDP growth that does not pay close attention to human capital may have reached redundancy many decades ago, such a model can be said to exist in the minds of many education policymakers even today. To illustrate, in South Africa the argument that better education cannot tackle poverty and that policies dealing with industry and the labour market need to accomplish this, is not an argument that one would easily find in the academic literature. Yet it is an argument that one can hear being made even by high-ranking government policymakers from time to time. Solow's industry-focussed model is thus by no means dead. Of course it is beyond argument that industry policy is important for growth. But why would anyone deny that human capital development has a vital role to play? One explanation would be an overly simplistic adoption of a Marxist framework where, as was discussed above, human capital development is not an important element. Another would be that the transmission between better education and higher income, for the individual and the country, has not been understood. Specifically, the link between human capital and productivity in the workplace may not be clear. Perhaps it should not surprise us that this link is not always understood as what it actually means, in terms of actual human behaviour, is often not clear in the literature. According to Boissiere (2004), this topic has received attention in the economic literature, though much of it has focussed on one occupation, namely farming.

Scepticism about the productivity-enhancing power of better education is likely to be driven by lingering notions that screening plays a large role in the labour market. Roughly, screening theory says that education serves as a sorter of workers in the labour market. The individual therefore acquires more education not because this makes her more productive, but because there is a general move towards more education and hence in order to maintain one's place in the queue, one is forced to acquire more education than one otherwise might (Spence, 1973; Stigitz, 1975). In his review of the literature, Boissiere (2004: 5) concludes that the evidence suggests that whilst screening is not non-existent, its impact in the labour market is relatively small.

2.3 The augmented Solow model

Mankiw, Romer and Weil (1992) (henceforth MRW) offer a persuasive and empirically supported defence of Solow's model. (David Romer of MRW is not the same as Paul Romer mentioned in the previous sub-section, and neither are the two related.) MRW do not respond directly to Romer's (1986) challenge, but they do present a robust empirical rendition of what they call the 'augmented Solow model'.

The theoretical version of this model is dealt with here, and the empirical version is one of the three empirical models examined in depth in section 4. The augmentation consists of adding a human capital variable H to equation (2), resulting in the following:

$$Y = AK^{\alpha}H^{\beta}L^{1-\alpha-\beta} \tag{6}$$

H has the same saving and depreciation dynamics as K and as reflected in equation (3). Presumably (this is not made explicit in MRW) saving occurs through investments in education (to a large degree the cost is an opportunity cost given the time intensity of education) and depreciation occurs when people become too old to work, or die.

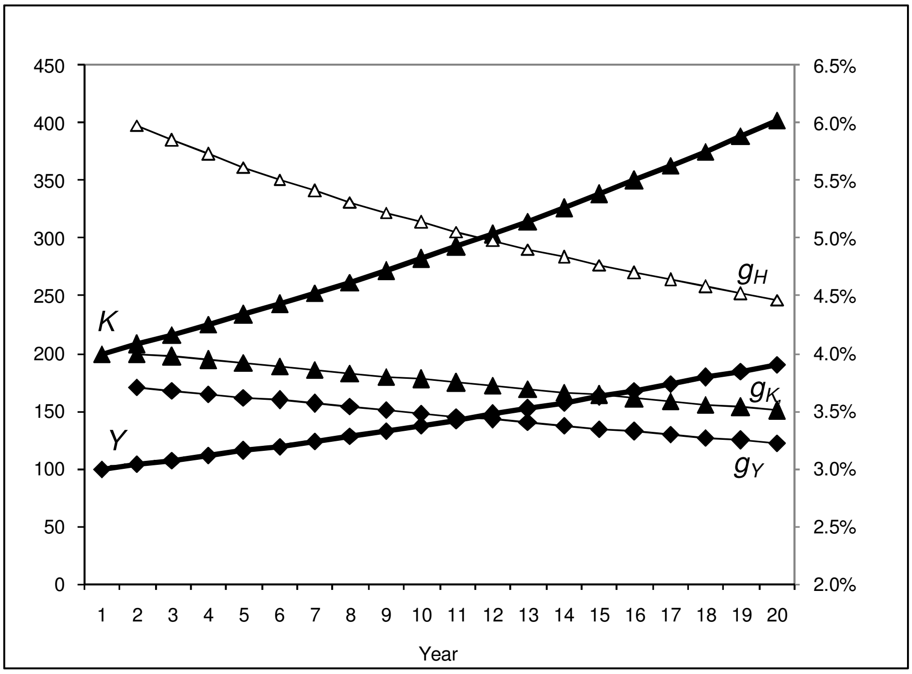
Equation (6) is Aghion and Howitt's (2009: 289) rendition of MRW's model. The original in MRW (1992: 416) is slightly different and appears as follows.

$$Y = K^{\alpha}H^{\beta}(AL)^{1-\alpha-\beta} \tag{7}$$

The two versions are not fundamentally different. Equation (7) implies considering effective labour AL and not raw labour L as a factor of production. Moreover, because the exponent of A in equation (6) is 1 whilst in equation (7) it is always less than 1, the first equation will always yield a higher value of Y and also greater growth in Y over time if the four right-hand variables and α and β are equal across the two equations.

The next graph illustrates the simulation of Figure 1 with a very small quantity of human capital H added in order to examine what the marginal impact of H might be. The starting values for the variables in equation (2) remained the same. The starting value for H is 1 and in order not to change the Solow model patterns too much β (the exponent for H) was set at 0.04, s for human capital was set at 0.0006 and δ for human capital at 0.0002.

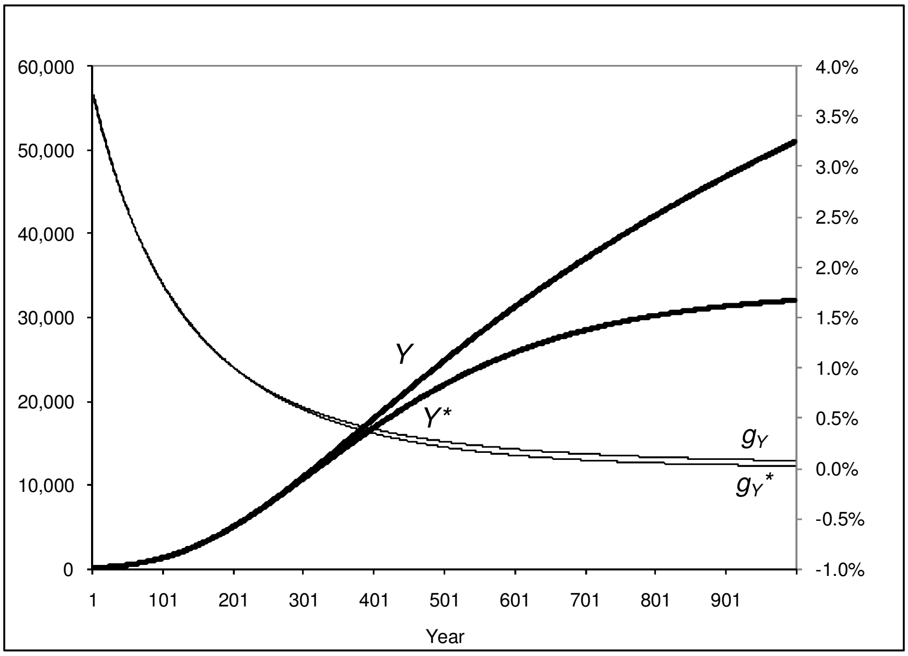
Figure 3: MRW's augmented Solow model



The insertion of the human capital variable results in a higher growth rate. Where growth declined to 2.8% in year 20 in the un-augmented Solow model, it declines only to 3.2% in the augmented model. As a result, Y in year 20 is now higher. Importantly, growth is enhanced not only by the direct effect of H on Y, but also through an indirect effect whereby the direct effect allows for more income to be invested in K. Very importantly, for H to have an effect on growth, H itself must grow. If human capital accumulation does not exceed human capital depreciation, then human capital does not enhance growth. In education terms, it is not enough for the country to be educated, the educational level of workers must improve over time if education is to have an impact on growth. This is why the augmented Solow model of MRW is said to belong to a class of models emphasising human capital *accumulation*, and not the *stock* of human capital (Aghion and Howitt, 2009: 287).

The long run version of Figure 3, given below, illustrates an important defining feature of the model. Eventually, a steady state level of Y is reached, but with human capital in the model this is higher and it takes longer to get there (compared to a similar model without human capital). Y^* and g_Y^* are from the un-augmented Solow model of Figure 2. Even without changes over time in the proportion of income devoted to human capital investment, differences in the human capital starting values between countries are, according to MRW, likely to cause different future steady states, and work against convergence between countries. That the inclusion of human capital should make the original Solow model much more compatible with non-convergence between countries, and hence vindicate the basic structure of the model, is a key argument in MRW.

Figure 4: MRW's augmented Solow model in the long run



The basic notion that levels of human capital and its growth over time should influence income and its growth to the degree that human capital deserves separate treatment in a growth model, would make sense to most education policymakers. Non-convergence, on the other hand, is less likely to be accepted as a feature of development by policymakers who consider the strong growth of certain Asian countries in the three decades between 1980 and 2010. The success of these countries has served to inspire hope that all developing countries can eventually become high income countries, and that the world can converge towards more or less the same high level of human welfare. To illustrate the pattern, between 1980 and 2010 Taiwan quadrupled its GDP per capita in real terms and surpassed Italy, in purchasing power parity terms. China, which in 1980 had an income per capita around one-fifth of that of South Africa, had more or less caught up to South Africa by 2010 (own analysis, using Heston, Summers and Aten, 2011). Even the general pattern has been weakly in the direction of convergence if one considers that the quintile of poorest countries in 1980 had on average a GDP per capita that was 3% of that of the richest quintile of countries, whilst by 2010 this percentage, using the same two groups of countries, had risen to 6%. However, if one does not use the same groups of countries for each year, and simply compares the average income of the bottom quintile to the top quintile, using all countries with non-missing data in both years, a less hopeful picture emerges. The percentage in fact drops from 3% to 2%. There is thus not unambiguous evidence supporting convergence over the last few decades. It is should be remembered that MRW's augmented Solow model does not imply that convergence cannot occur, but rather that non-convergence may occur as a result of different initial levels of human capital, or different human capital accumulation rates. The notion that not all countries will eventually converge to exactly the same level of per capita income seems uncontroversial.

2.4 The AK model

The AK model was an early attempt to explain non-convergence and non-diminishing returns to capital investment in per capita terms. It started as an entirely theoretical exercise centred to a large extent around articles by Frankel (1962) and Arrow (1962). The AK model uses as its point of departure the following Harrod-Domar form (Frankel, 1962: 995):

$$Y = AK \tag{8}$$

The model takes its name from this basic form, which can be seen as a special case of the Cobb-Douglas form with the exponent for *K* set to 1 and the exponent for *L* set to 0 (making *L* equal 1). Clearly, if *A* is constant then growth in *Y* will be proportional to growth in *K*, meaning constant returns to scale, and if *A* grows there will be increasing returns to *K*. Diminishing returns to *K* are impossible. This puts the model firmly outside the neoclassical class of models, which consider diminishing returns to be indispensible.

The AK model is said to be endogenous as it provides an explanation, albeit a very simple one, for growth of *A*. Physical capital generates positive externalities in the form of improved knowledge and skills in society, so *A* is a positive function of *K* (Aghion and Howitt, 2009: 50). The suggestion is that economies of scale are important. A larger country would have a higher value of *A* than a smaller country with the same amount of *K* per capita. *A* is thus barely endogenised, given the obvious complexity of *A*. As Aghion and Howitt (2009: 323) put it: 'There are many channels through which societies accumulate knowledge, including formal education, on-the-job training, basic scientific research, learning by doing, process innovations, and product innovations.'

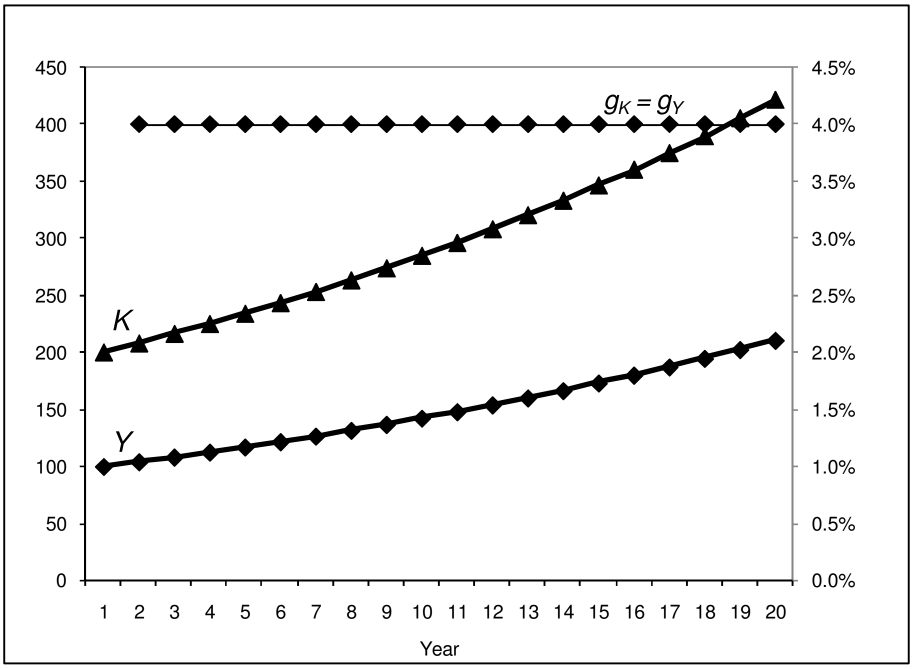
In one sense the AK model is thoroughly neoclassical. It assumes that firms are competitive and profit-maximising and that publicly accessible knowledge (in many senses *A*) grows due to positive and unintended externalities in the capital accumulation process. The model does not see a conflict between competition and innovation in the way that the Schumpeterian models examined below in sections 2.7 to 2.9 do.

The AK model combines equations (3) and (8) to obtain the following equation for growth:

$$g_Y = sA - \delta \tag{9}$$

The next graph illustrates the AK model. In this simulation, *A* is kept constant. It was critical to set this variable at a value which did not result in implausible trends. The other variables were all given the same starting values as in the Figure 1 simulation of Solow's neoclassical model.

Figure 5: The AK model



The main difference between this outcome and that of the Figure 1 simulation is that growth is constant here, implying no movement towards a steady state of *Y*. The most developed countries will always remain ahead in the development trajectory.

If the savings rate *s* is reduced to 0.15, a result similar to that for the neoclassical model is obtained. Growth in *Y* diminishes, in the case of the AK model to 1.5% (though it remains constant). Interestingly, if we simulate once again a country with a low saving rate (*s* is 0.15) and a low initial capital endowment (capital is 150), then g_Y does not fall. Instead, it remains at the 4% level seen in Figure 5. But this adjustment requires one to increase *A* to balance the equation. The model therefore forces one to compensate for lower capital intensity with higher productivity *A* (the starting value for *Y* has not been adjusted).

The idea that physical capital should in itself stimulate human capital, as implied by the AK model, is an idea that education policymakers often consider. Specifically, in a context of weak teachers, policymakers often turn to the possibility that computers, computer software, textbooks and science equipment could in some way compensate for the poor human capital of the teachers, in particular if the latter is such a serious problem that the scope for in-service training solutions is severely limited. Of course the AK model is not making this claim specifically for the education system, but rather in terms of the labour market and the economy as a whole (which would include the education system).

2.5 The Lucas model

Lucas's (1988) model is entirely theoretical, though it is informed by Denison's (1974) empirical work on the growth of the US economy using the Solow model. Aghion and Howitt (2009: 287) consider Lucas's model to be an AK model due to its non-convergence structure. They also consider it to be one of two models emphasising human capital accumulation, as opposed to the stock of human capital (the other model is the MRW model discussed above).

Despite its AK qualities, the Lucas model does assume diminishing returns to *K*. The exponent β in the following is always less than 1:

$$Y = K^{\beta}(uH)^{1-\beta} \tag{10}$$

The variable *u* represents the proportion of a person's (or all people's) time spent working, and not studying. Income *Y* is thus explicitly a function of both physical capital *K* and that part of human capital *H* not invested in creating further human capital. Lucas's equation reflects his insistence that macroeconomic models need microeconomic foundations (in this case a basis in the education investment decisions of households and individuals), and his interest in incorporating elements of human capital theory into a growth model (Lucas, 1988: 15). Lucas's full rendition of the equation includes an additional element, an external effect of *H*, or an income-enhancing effect that is not perceived by the individual but is felt by the economy as a whole (Lucas, 1988: 18). Put differently, there is a positive externality arising out of the individual's investment in education. Here this element is ignored for the sake of simplicity.

In the Lucas model physical capital is accumulated according to the following equation, which excludes depreciation (represented in the earlier equation (3) by δK):

$$\dot{K} = Y - C \tag{11}$$

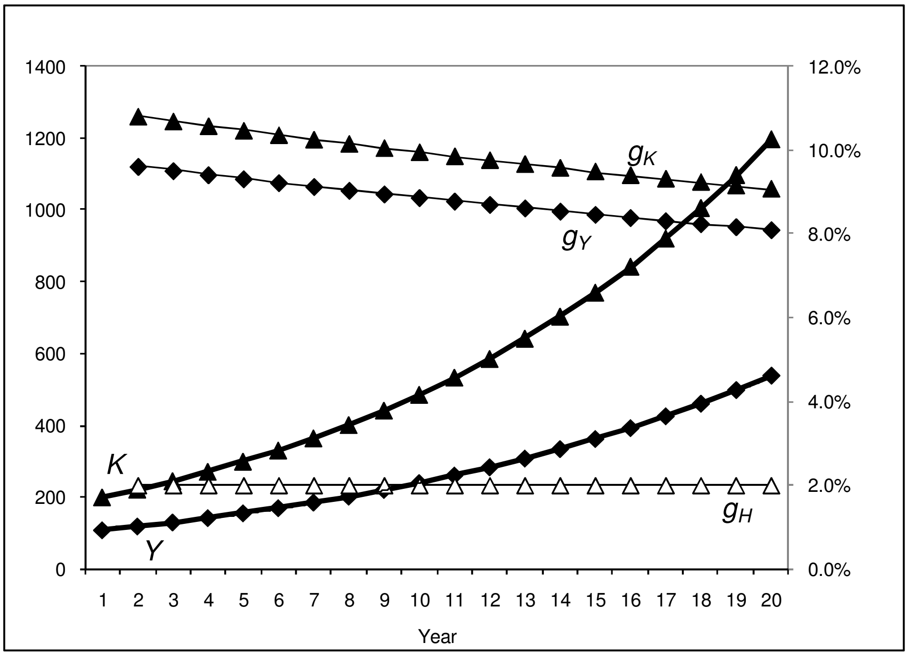
C is aggregate consumption. This is of course just a different way of saying that the capital stock grows by the amount saved, as in the neoclassical model. The next equation explains how *H* changes across two periods of time.

$$\dot{H} = \gamma H(1-u) \tag{12}$$

Here γ refers to the productivity of the education system. Growth in human capital is thus higher the more time is spent in education, and the more productive the schools are. The above implies that growth in *H* will be constant as long as the parameters γ and *u* remain the same.

The combination of the previous three equations allows for the simulation illustrated in Figure 6. Whilst in the AK model simulations discussed above growth of *Y* and *K* is always constant, plausible manipulation of the values in the Lucas model allows growth of *Y* and *K* to increase, decrease or remain constant. In Figure 6 decreasing values of g_Y and g_K are simulated so that scenarios comparable to those in Figure 1 and Figure 3, where these two variables also reduce over time, are produced.

Figure 6: The Lucas model



The above scenario uses the following parameters: β is 0.87; u is 0.9; consumption C is always 0.8 of income Y; γ is 0.2. Maintaining a constant ratio of consumption to income results in a constant value for g_H.

Where the AK model sees improvements in human capital, in the sense of improvements in A, as a function of the interaction between physical and human capital, the Lucas model discussed above presents a notion of human capital growth that would be more familiar to education planners. Time is a core ingredient in the education process and it is not easily substituted in the education production process. Whilst in the production of, say, cathedrals, it has been possible to reduce the time input through better machinery, no such substitution seems possible in education, at least not in any comparable sense. However, though there may be constraints insofar as it appears difficult, for instance, for the average child to become literate and numerate in less than around six years, within this constraint there is a large range with respect to the possible depth and quality of the literacy outcomes of a schooling system. Even similarly resourced schooling systems with apparently similar organisational structures can be significantly different with respect to their educational outcomes. Lucas's acknowledgement of this through the internal efficiency parameter γ is thus important.

Within economics, Lucas can be said to be a proponent of what educationists refer to as lifelong learning and economists 'learning by doing'. As Kenneth Arrow, Nobel laureate for his work on general equilibrium theory and the economist generally considered to have introduced 'learning by doing' into the economics lexicon, puts it (Arrow, 1962: 155):

> Learning is the product of experience. Learning can only take place through the attempt to solve a problem and therefore only takes place during activity. Even the Gestalt and other field theorists, who stress the role of insight in the solution of problems ... have to assign a significant role to previous experiences in modifying the individual's perception.

This has two implications for human capital accumulation. On the one hand, it points to the importance of having a sufficient degree of experiential learning and problem-solving activities within the formal education system. On the other, learning outside of formal education contexts needs to be acknowledged as an important contributor towards human capital accumulation, in modelling growth and formulating policy on productivity. These are ideas virtually all education policymakers embrace. Lucas's emphasis on time in the sense of the 1 minus *u* seen in equation (12) should be seen in this context. Sufficient time must be devoted to formal education, but organising the workplace in a manner that maximises on-the-job learning, in effect blurring the distinction between work time and learning time, is important too.

2.6 Product variety model

The product variety model is described by Aghion and Howitt (2009: 14) as part of a second wave of endogenous and post-neoclassical models incorporating considerable attention to innovation, and explaining the determination of *A* in a more detailed manner than in the AK model. They see the product variety model as the first branch of this wave, the second being the Schumpeterian models described below.

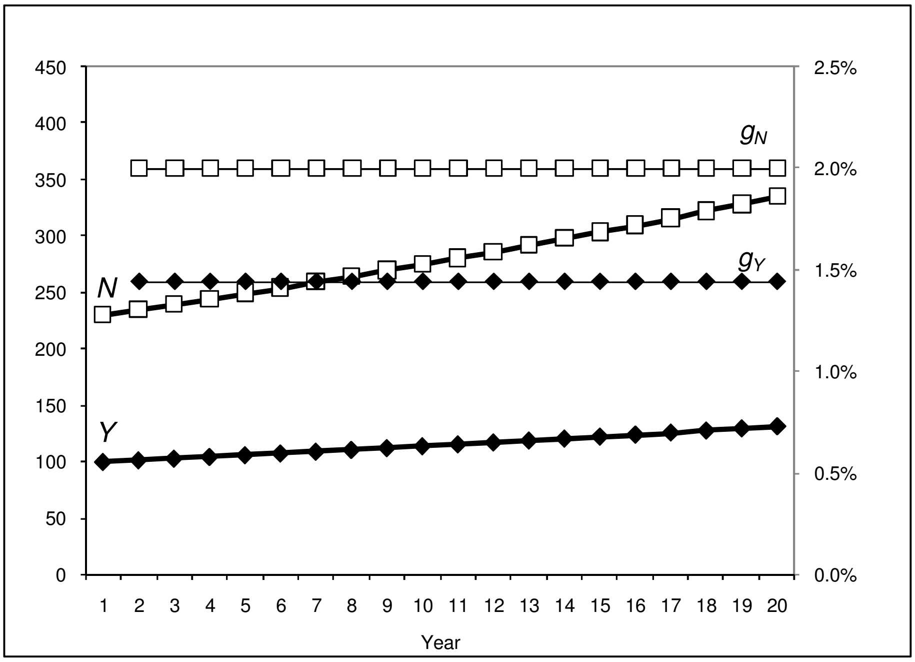
What Aghion and Howitt (2009) call the product variety model stems from the theoretical model of Dixit and Stiglitz (1977) titled 'Monopolistic competition and optimum product diversity'. The most basic form of the model is as follows (Aghion and Howitt, 2009: 15):

$$Y = N^{1-\alpha} K^{\alpha} \tag{13}$$

The structure resembles a Cobb-Douglas one. *N* is the number of industries in the country. In the simplest form of the model, *K* is exogenous, meaning that in the simulation, values in each year can be calculated without reference to values in the previous year. The starting value for *N* was set at 230 (the number of major industry categories used by Statistics South Africa, based on ISIC[7]), and this was the only exogenous variable that was allowed to increase (it increased by 2% per year). It is thus assumed that innovation creates entirely new industries. The capital/output ratio in year 1 was set at 2.0. The value of α was adjusted to create a plausible scenario, which is illustrated below.

[7] United Nations International Standard Industrial Classification of All Economic Activities. See United Nations Statistics Division (2009b).

Figure 7: Product variety model



The fact that the simulation allowed for non-integer values of *N* obviously makes the pattern smoother than it would be in a more realistic simulation. The product variety model assumes that the greater the number of industries, and hence the more diversified the economy, the more the innovation, and this is what stimulates growth. The simulation in Figure 7 has constant growth in *N,* and constant growth in *Y* at just under 1.5%. The explanation for this lower rate of growth in *Y* (compared to the 2.0% for *N*) reveals an important point about the model. Because *K* is kept at a constant value of 200 in the simulation, the same amount *K* is being spread ever more thinly across a greater number of industries *N*. The declining capital stock within each industry offsets some of the growth effects of greater diversity. If, on the other hand, *K* increases at 2% per year (like *N*), then *Y* will also grow at 2%.

The product variety model implies that industry policy should encourage diversity, and that education policies should nurture new technology specialisations. A key question in the current industrial development debates as reflected in, for instance, Lin and Chang (2012), is the degree to which the state, in increasing *N* in equation (13), should use temporary protectionism to nurture new industries (a comparative-advantage-defying approach), as opposed to just creating an enabling environment for competitive industries (a comparative-advantage-following approach). The education policymaker's segment of this important policy question is the degree to which training in new technologies is possible in the absence of that technology within industry. For instance, if the use of nanotechnology is to be stimulated within the economy, to what extent is training in this technology dependent on the existence of at least some presence of this technology in some industrial sector? This presence could have been introduced through a government programme that incentivised foreign investment and the importing of foreign skills. To put the question differently, how do complementarities between the education system and industry work when new technologies grow within a country? This question does not seem to have received significant attention in the academic literature, perhaps because measuring these complementarities would be difficult. At least on a theoretical level, one can speculate that the presence of the new technology in industry would improve the quality of training in this technology within the education sector through two channels. Firstly, education institutions could draw from the human expertise existing within the country and, secondly, students would be incentivised by the existence of concrete opportunities of employment within the country in the new field.

2.7 Schumpeterian model of Nelson and Phelps

Three Schumpeterian models are discussed in this sub-section and the subsequent two. The Nelson and Phelps (1966) model, an early growth with education model often referred to in the subsequent literature, is considered an early Schumpeterian model by Aghion and Howitt (2009: 298).

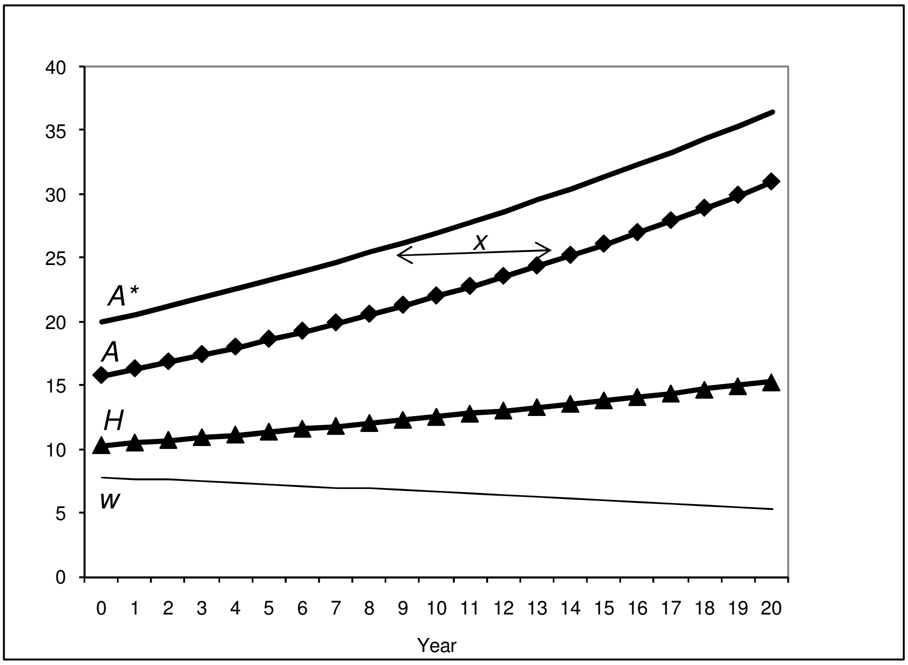
Some background on the work of Joseph Schumpeter (1883-1950), an Austrian economist, seems in place. Schumpeter, who used an exclusively narrative and non-mathematical style of argument, views capitalism as an economic system harbouring the seeds of its own destruction, and predicts high levels of monopolisation. In this respect he is similar to Marx, but this is as far as the similarity goes. For Schumpeter, society's support for capitalism rests largely on the opportunities afforded to entrepreneurs, who innovate in the production process. He distinguishes innovation from invention, the latter being more technical, and the former more economic and organisational. Innovation involves continual 'creative destruction', a term Schumpeter is often associated with. Innovations result in the success of innovating firms and the demise of firms that do not innovate successfully. Over time firms learn this is the pattern, and increasingly they try to organise innovation rationally, through research institutions, which become increasingly bureaucratised. Successful firms are increasingly large firms, and monopolies begin to dominate the market. A situation arises in which 'innovation itself is being subjected to routine' in order to 'make technical change a predictable process' and this is what leads to dissatisfaction amongst more traditional entrepreneurs and 'an army of critical and frustrated intellectuals', who then bring about the demise of this highly monopolistic form of capitalism (Heertje, 2008: 321). One presumes that this demise would lead to the re-emergence of a more competitive capitalism, and a repeat of the whole cycle, but this does not seem to be clear in Schumpeter's work. Schumpeter's version of historical determinism has been criticised, but his view of innovation and imitation as two key organising phenomena within capitalism has been embraced, for instance by Aghion and Howitt (2009).

The simpler of two models put forward by Nelson and Phelps (1966: 72) is discussed here. (Edmund Phelps is a Nobel economics laureate for his work on inter-temporal tradeoffs in macroeconomic policy.) In this model growth in Y is considered to be synonymous with growth in A, as in the subsequent two Schumpeterian models. Human capital H is a determinant of A (and Y), but unlike the MRW and Lucas models, the Nelson and Phelps model emphasises within its basic form the stock of human capital, as opposed to its accumulation, in the growth process. The basic form is as follows:

$$A_t = A^*_{t=0} e^{\lambda(t-w(H))} \quad (14)$$

A, called the 'index of technology' in Nelson and Phelps (1966), is specific to a country, as in A_t above, but there is also a 'global technology frontier' A^*, or what was referred to earlier as the currently known best way of doing things. At any point in time, *A* is less than A^* as there is a perpetual lag whereby countries take some time to adopt the best available technologies, and what is best is continually advancing. Conceivably, all countries could experience this lag as even the most advanced countries would have reached the frontier in some industries, but not in others. In equation (14) *A* at current time *t* is equal to A^*, or the global technology frontier, at time 0, multiplied by the base of the natural log *e* to the power of λ, a constant determining how fast A^* grows, multiplied by current time *t* minus a lag *w* which is a decreasing function of human capital *H*. The lag is thus shorter the more educated the country's human capital is. For a country to decrease the lag, it must increase its stock of human capital *H*. The next graph illustrates the case of a country that is continually increasing its stock of *H*. Though not clearly visible in the graph, the horizontal distance *x* is diminishing over time due to the increase in *H*. Logically, whilst the gap is being closed, growth in *A* will exceed growth in A^*, but when *H* stabilises, the country will simply follow the growth of A^*.

Figure 8: Nelson and Phelps



2.8 Schumpeterian model with creative destruction

This and the next Schumpeterian model both stem from work by Phillipe Aghion, a French economist. The model discussed here is a relatively simple one from Aghion and Howitt (2009: 15) which does not include a human capital variable (the next one includes several). The basic form resembles the Cobb-Douglas form of equation (1):

$$Y = A^{1-\alpha}K^{\alpha} \tag{15}$$

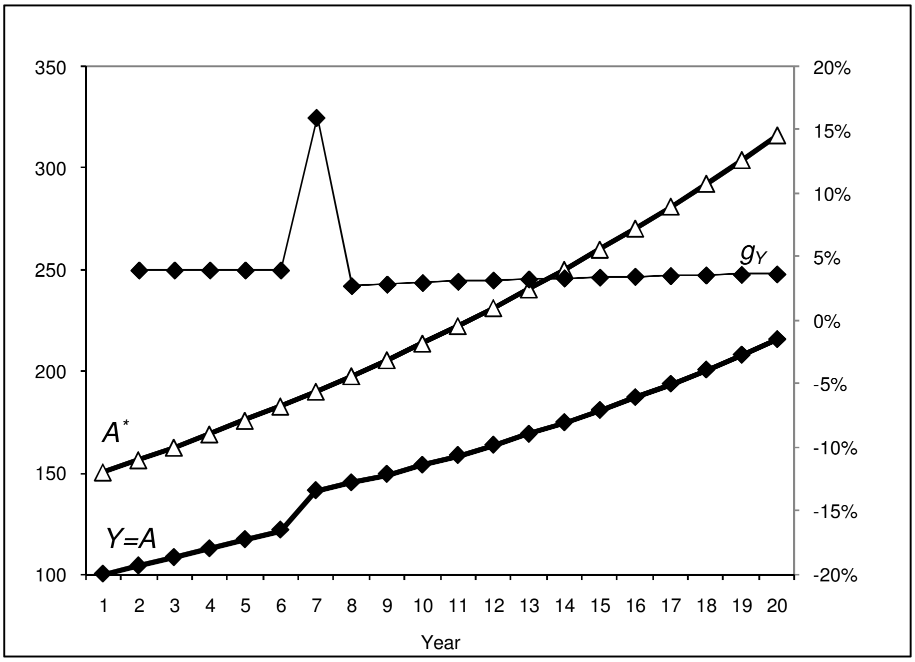
Here, however, *A* carries an exponent less than 1, assuming α>0, meaning there are diminishing returns to *A* (there are also diminishing returns to *K*). The creative destruction element of this model appears in the determination of *A*:

$$A_t = A_{t-1} + \mu_n(\gamma - 1)A_{t-1} + \mu_m\left(A^*_{t-1} - A_{t-1}\right) \quad (16)$$

To simplify the discussion and the simulation, K is assumed to be constant (the emphasis in the model is on the determination of A, not of K). Again A^* is the global technology frontier. In the full specification of the model, A^* is the global technology frontier for one industry and equation (16) refers to the dynamics of one industry. In the interests of simplicity the industry subscripts are excluded from equation (16). Growth of Y in the country is the aggregation of the industry-specific growth. The variables μ_n and μ_m are the frequency of innovation and imitation. This is best understood through reference to the country as a whole. If there are 20 industries, for instance, and we expect there to be a successful imitation of a technology from elsewhere occurring in one industry in each year, we can think of μ_m as 0.05 (1 over 20). Similarly, if μ_n is 0.05 this could mean that in any year 1 of 20 industries advances its technology to a point beyond the global technology frontier for that industry. How much further than the frontier is indicated by the factor γ, which must logically be greater than 1 and is also indicative of the productivity enhancement achieved as a result of the innovation. The innovation would result in an increase in not just A but also A^*. Imitation on the other hand, changes just A of the country. The global A^* remains unchanged.

There are several ways in which one could represent this Schumpeterian model graphically. In Figure 9 below μ_m is considered to be a constant 0.08 every year, whilst μ_n is considered to be zero in all years, except for year t=7 when it becomes 0.03 so that an innovation (in at least one of the country's industries) is forced to occur. γ is set at 5.0, the starting point for A is 100, and that for A^* 150. A^* is assumed to increase at an exogenous rate of 4% per year, which would be caused by countries jumping the global technology frontier in specific industries on a continuous basis. In order to obtain a starting value for Y of 100 and for K of 200 (K will not change), in line with previous simulations, we must adjust the exponent α from equation (15). This adjustment forces α to become almost zero indicating that in our simulation Y is virtually equal to A.

Figure 9: Schumpeterian with creative destruction



There is a spike in the growth of Y in year 7, from 4% to 16% and then down to 2.8%. The increase in g_Y occurs due to the improved productivity brought about by the innovation, which moves the country abruptly to a new level of income Y. The fact that there is some growth even in other years is due to the fact that there is new imitation being initiated in each year. If there had been a year with no imitation or innovation (with both μ_n and μ_m equal to zero in other words), there would have been no growth in that year.

Very importantly, after the innovation in year 7, growth declines to a level that is lower than it was before year 7 – it drops to 2.8% in year 8. This is because the gap between A and A^* has been narrowed, and with constant μ_m equation (16) tells us that the growth in A due to imitation would have to decrease with a reduction in the ratio of A^* to A – in our simulation this ratio decreases from 1.50 in year 6 to 1.35 in year 7. However, in the absence of further innovations in our country the ratio of A to A^* rises again to 1.47 by year 20, and consequently g_Y increases from 2.8% to 3.7% from year 8 to year 20. Basically continual innovation elsewhere has introduced more room for local imitation. The fact that growth is a positive function of the distance to the global technology frontier, measured as the ratio of A^* to A, can be seen in the following equation for g_Y:

$$g_Y = \mu_n(\gamma - 1) + \mu_m\left(\frac{A*}{A} - 1\right) \qquad (17)$$

The full version of this Schumpeterian model explains μ_n in microeconomic terms (Aghion and Howitt, 2009: 88).

$$\mu_n = \lambda\left(\frac{R}{A*}\right)^{\sigma} \qquad (18)$$

Here R is expenditure on research. The higher the productivity of the research sector, or the higher λ is, the greater the probability of an innovation. And the greater the investment in research R by firms, the greater this probability. However, there is an important relationship between R and the global technology frontier A^*. The more advanced the frontier, the smaller the returns to investment in research R. In other words, there are diminishing returns to research investments over time. Investing resources in innovation research in the year 1900 was likely to yield better returns than investing the same resources in 2000, by which time knowledge advancements would have made research more complex and costly. Finally, σ represents a Cobb-Douglas type exponent representing the expenditure elasticity of the level of research outputs.

2.9 Schumpeterian model with two levels of human capital

The final Schumpeterian model examined here is one put forward by Vandenbussche, Aghion and Meghir (2006) (henceforth VAM). The empirical version of VAM's model is discussed in Part II of the dissertation. As in the previous model, the focus is at the industry level, and the country picture consists of the aggregation of all industries. Again, for the sake of simplicity, industry subscripts are dropped. The basic form of the model is again a Cobb-Douglas one.

$$Y = A^{1-\alpha} x^{\alpha} \tag{19}$$

Instead of using capital stock K as a determinant of growth, this model uses intermediate product x. This latter variable could, for instance, refer to the plastics used to produce a computer. The computers produced, and encapsulated within income Y, would be a function of the value of the plastic used x, the productivity of the manufacturing process A and the Cobb-Douglas exponents $1-\alpha$ and α. The following equation demonstrates how A would be determined.

$$A_t = A_{t-1} + \lambda\left(\gamma u_n^{\phi} s_n^{1-\phi} A_{t-1} + u_m^{\sigma} s_m^{1-\sigma}\left(A_{t-1}^{*} - A_{t-1}\right)\right) \tag{20}$$

This equation is similar to equation (16), but there are some key differences. There are Cobb-Douglas forms for innovation (n) and imitation (m). For instance, innovation has two inputs with exponents adding up to 1. The two inputs are the amount of unskilled labour u_n and the amount of skilled labour s_n dedicated towards innovation activities. The structure is similar for imitation activities. One may wonder why unskilled labour would be dedicated towards innovation. This is not explained by VAM, but the value of u_n can be zero, and 'unskilled' is somewhat of a misnomer as VAM regard workers without post-secondary education as unskilled. λ refers to the efficiency of the overall process of technological development, γ refers to the efficiency of innovation relative to that of imitation, and σ and φ are the familiar Cobb-Douglas exponents. The model borrows from the Rybczynski model of international trade, the difference being that here different countries are converted to two different sectors of the economy, imitation and innovation. The total number of unskilled workers U and the total number of skilled workers S are exogenous (and would largely be an outcome of the country's education policies), but the allocation of the two types of workers across the imitation and innovation sectors is endogenously determined according to the following two formulas:

$$s_m = \frac{U\left(\frac{(1-\sigma)\psi^{\sigma}(1-a)}{(1-\phi)\gamma a}\right)^{\frac{1}{\sigma-\phi}} - S}{\psi - 1} \tag{21}$$

$$u_m = \frac{\psi U s_m}{S + (\psi - 1)s_m} \tag{22}$$

Here a represents the ratio A / A^* where 1 is the maximum and would mean the country had caught up to the global technology frontier (the ratio referred to here is the inverse of the ratio referred to in (17)). ψ is calculated as follows:

$$\psi = \frac{\sigma(1-\phi)}{(1-\sigma)\phi} \tag{23}$$

The essence of the previous three equations is that the allocation of skilled and unskilled labour across the two sectors depends on, firstly, the sizes of the two pools of labour and, secondly, the distance of the country from the global technology frontier (the details of these dynamics are explained below).

The VAM model includes the determination of the amount of intermediate products x in equation (19) according to equation (24) below. x is a function of A in the sense that the better the technology and productivity A in the country (or industry), the more x is produced. In the complete dissection of equation (24), which is not explained here, monopoly firms are a necessity as competitive firms would not have a large enough incentive to invest sufficiently in research as they would not enjoy monopoly profits following the successful innovation.

$$x = \alpha^{2/1-\alpha} A \tag{24}$$

This Schumpeterian model was first simulated below with just 13% of all workers assumed to be skilled across all years. Figure 10 indicates that the proportion of all skilled workers engaged in the innovation sector would rise, from around 40% to around 55%. Though this is not clearly visible in the graph, the proportion of unskilled workers in the innovation sector would, whilst remaining low, also increase slightly. This is because imitation activities are continuously improving the closeness to the global technology frontier, or a, which in turn increases the marginal returns to both types of labour in the innovation sector. Put differently, the more advanced the country is technologically, the more likely it is that firms in the country will be able to add something new to the world's stock of innovations. As a result of a mix of imitation and innovation, aggregate income Y (or A) expands by 53% altogether over the 20 years, or 2.1% a year[8]. This is not an impressive growth rate, but it should be remembered that over the period there has been no human capital improvement in the sense of the percentage of workers who are skilled.

[8] In simulating the model, it was necessary to use a from the previous year in equation (21) in order to avoid a circular reference problem in Excel.

Figure 10: Schumpeterian with two levels of HC (I)

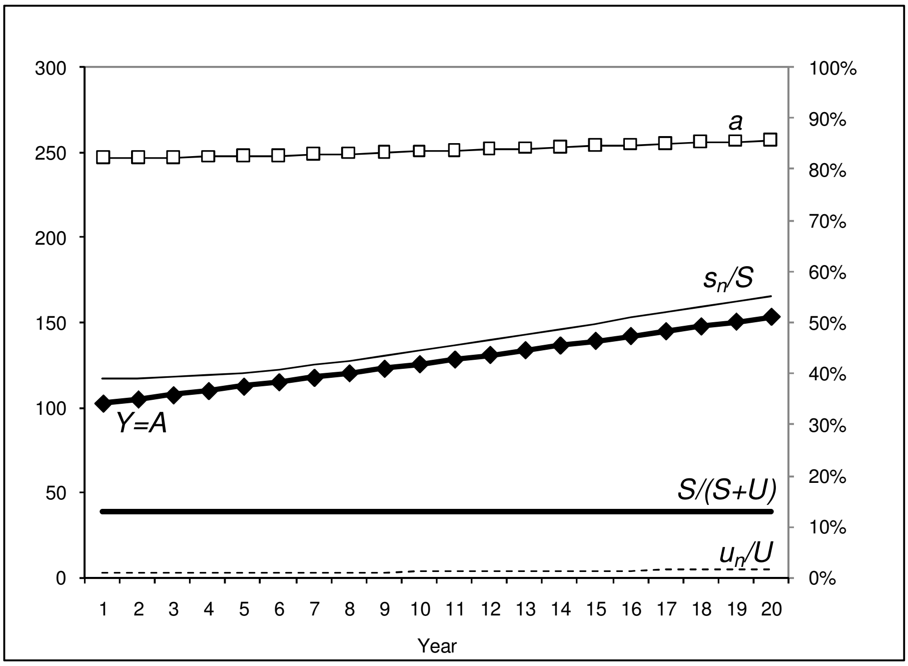
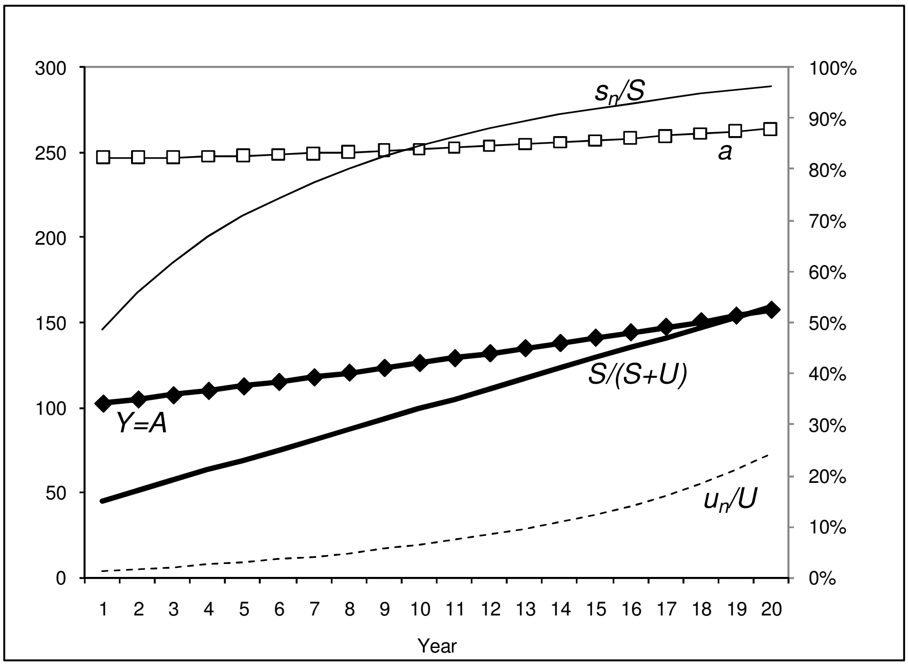


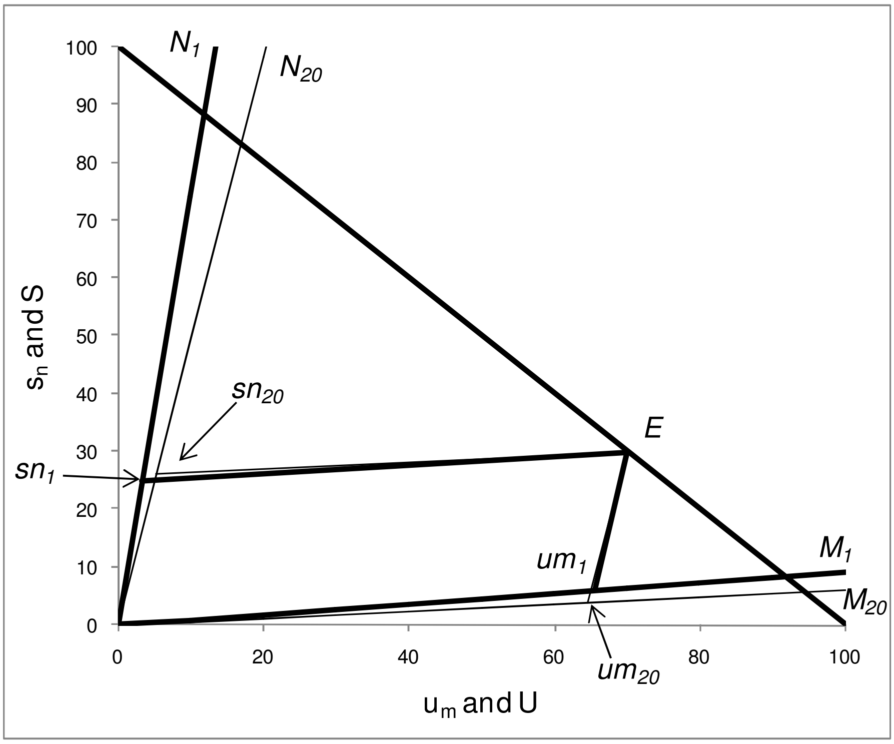
Figure 11 below illustrates an alternative simulation where the proportion of workers who are skilled does increase exogenously from 13% to 53% over the 20 years as a result of education policy interventions. As will be seen, this very large degree of human capital improvement seems to be necessary for even modest gains in growth to be realised. The shift has the effect of improving the outputs of both the imitation and innovation sectors, relative to the previous scenario. This results in a somewhat higher level of *a* in year 20. Put differently, the distance to the technology frontier has been narrowed to a greater degree than in the previous scenario. Growth in *Y* improves. The growth in *Y* over 20 years is now 57% (against 53% in the previous simulation). This seems a rather limited reward for the very large increase in the proportion of skilled workers, but it seemed difficult to adjust the model to simulate higher growth. With respect to the allocation of workers across the two sectors, the movement of both skilled and unskilled workers towards innovation is much larger than before. In fact, the simulation indicates that by year 30 (off the graph) 100% of skilled and unskilled workers would be engaged in innovation.

Figure 11: Schumpeterian with two levels of HC (II)



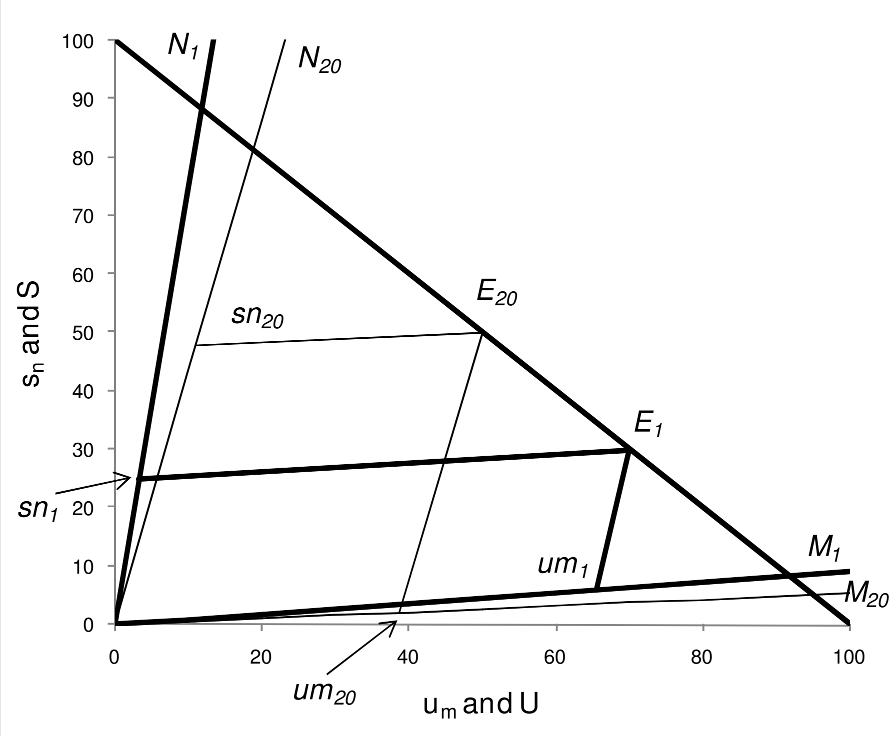
VAM summarise the dynamics of their model in a graph that is reproduced below as Figure 12, using more or less the values from the two simulations presented above. Figure 12 illustrates what happens when the percentage of skilled workers remains constant, but instead of the constant level of 13% seen in Figure 10, a level of 30% is used as this produces a clearer picture. The problem with the 30% level, however, and the reason why a lower level was used earlier, is that 30% results in a situation where the majority of both skilled and unskilled workers are in the innovation sector. It seems unrealistic for such a low level of education (30% of workers are skilled) to be associated with an economy that innovates more than it imitates. This could be indicative of a limited real world relevance of the model. In Figure 12, point *E* indicates the total initial number of skilled and unskilled workers. Of a total of 100 workers, 30 are skilled (vertical axis) and 70 are unskilled (horizontal axis). The thick-lined parallelogram between the origin of the graph and point *E* represents the situation in year 1. The slopes of lines N_1 and M_1 in that year are a function of the distance to the global technology frontier and this, plus the percentage of workers who are skilled, determine the number of skilled workers engaged in innovation, sn_1, and the number of unskilled workers engaged in imitation, um_1 (and by implication skilled workers in imitation and unskilled workers in innovation). For instance, of the 30 skilled workers in year 1, 25 are employed in innovation industries – read the top-left corner of the thick-lined parallelogram against the vertical axis. Thus 5 skilled workers must be working in the imitation sector. After 20 years, the economy has moved closer towards the global technology frontier, and consequently lines *N* and *M* have tilted clockwise. This results in a slightly higher number of skilled workers engaging in innovation, and in slightly fewer unskilled workers engaging in imitation – see the thin-lined parallelogram and its corners sn_{20} and um_{20}. The total number of workers has remained 100 and the number of skilled workers is still 30.

Figure 12: A swing towards innovation (I)



In Figure 13 we see what happens if over 20 years the proportion of the 100 workers who are skilled increases from 30 to 50. The change in the composition of the country's human capital shifts *E* from E_1 to E_{20}. The fact that by year 20 the country has moved closer to the global technology frontier (compared to the Figure 12 scenario) results in a more pronounced tilting of the *N* and *M* lines. This tilting results in a larger proportion of skilled workers being in innovation, and a lower proportion of unskilled workers being in imitation.

Figure 13: A swing towards innovation (II)



One thing that the three Schumpeterian models discussed above share, but not the models preceding those, is an explicit recognition that countries learn from each other. Indeed, cross-country learning is a key element of the development process according to these models. By implication, cross-country competition is important and one presumes that some of the Schumpeterian 'destruction' would take the form of countries that fail to compete successfully in a globalised economy. Competing is a matter of both innovation and imitation. The emphasis on imitating seems important, partly because imitation may appear to be a less noble pursuit than innovation, to the education policymaker and others. The Schumpeterian models imply imitation needs to be taken seriously, in particular in the case of a developing country. This means studying carefully what other countries do, assessing the transferability of certain technologies and accepting that one may do better by emulating someone else than developing a completely home-grown solution. One can think of the education system itself as a key industry that is subject to the dynamics of equation (16), in other words the need for the right mix of ongoing imitation and occasional innovation, where the latter carries benefits for other countries in the form of new knowledge and new recipes for improving the education system. The academic contribution towards cross-country learning in education policymaking should lie largely within the field of comparative education. To quote Psacharopoulos (1990: 371):

> Comparative education – or sheer history – can teach us a lot from different country experiences in expanding coverage and serving efficiency and equity goals.

As was discussed previously, Psacharopoulos (1990) believes comparative education could make a stronger contribution to cross-country learning in education, and hence educational improvement. The literature could be more practical and focus less on abstract and conceptual matters.

Globalisation, which is implicit in the Schumpeterian models, is considered a malevolent force by several educationists. As argued by Carnoy and Rhoten (2002), a deeper look at the impacts of globalisation on education reveal that much of the unease amongst analysts of the education sector (not so much education policymakers) can be attributed to certain misunderstandings of, for instance, the past impacts of globalisation. Global learning and even competitiveness can strengthen a country's education system, even though there are risks, such as insensitivity to local context within policy design, which are mostly avoidable if policymakers take the necessary precautions.

2.10 Conclusion

The overview of fifty years of growth theory presented above can only scratch the surface of the available literature. Virtually all of the concepts dealt with seem fundamental for an understanding of growth. The evolution of growth models has not been a matter of new concerns simply being added to previous concerns, or a matter of increasing complexity. There has been a 'creative destruction' of sorts as new concerns replaced older concerns. For instance, the concern with whether or not there are diminishing returns to physical capital has to a large extent given way to concerns around how human capital impacts on growth and how monopolies influence innovation. The inherent constraints of the modelling process require us to select only a few items from the 'menu' at a time. However, it is important to distinguish between the model itself, and the narrative about the model. It is possible to deal with a broader range of issues in the narrative than in the model itself. The narrative should partly be about discussing what did not fit into the model, and how this might affect the model's policy application.

Table 1 offers a structure that can assist the summing up. Differences in the interpretation of *A* across the various models serve to remind us of the importance of being clear on this matter. Concepts such as 'level of technology' and 'productivity' overlap, but they are not synonymous. Explaining the determination of *A,* however one defines this, is important for the modelling of growth. Where the macroeconomic model is grounded on microeconomic concepts of the firm, more specificity is introduced, which can assist in understanding precisely what policies promote growth. Just one of the models presented an endogenous explanation of human capital growth. This was the Lucas model, where both more time spent on education, including learning-by-doing at the workplace, in combination with efficiency improvements within the education system, lead to higher levels of human capital. The modelling of two levels of human capital, as in the final model by Vandenbussche, Aghion and Meghir (2006), also seems valuable, though here what actually contributes to human capital growth was not dealt with. Contrasting accumulation to stock effects when looking at human capital receives considerable attention in the literature. This is perhaps the most difficult aspect of the models to translate into practical terms. Both accumulation and the level of human capital, in the sense of the capabilities of the average person in the economy, are important in the real world and it is difficult to explain, from a practical policymaking perspective, why one would want to emphasise the one over the other. The theoretical nature of the models discussed, or their lack of concrete timeframes and historical contexts, makes it especially difficult to interpret the implications of the accumulation versus level debate. Lastly, the question of whether there is convergence towards a steady state of income is perhaps largely academic, especially if the implication of a convergence model is that this steady state is reached hundreds of years into the future. Yet that some non-rich countries develop faster than others is an inescapable reality and the basis for much enquiry, and a key point of departure for this dissertation. As Lucas (1988: 4-5) puts it:

> I do not see how one can look at figures like these [on differences in economic growth] without seeing them as representing *possibilities.* Is there some action a government of India could take that would lead the Indian economy to grow like Indonesia's or Egypt's? If so, *what* exactly? If not, what is it about the 'nature of India' that makes it so? The consequences for human welfare involved in questions like these are simply staggering: Once one starts to think about them, it is hard to think of anything else.

Table 1: Summary of model features

	Term used for *A*	Explains technology *A* endogenously	Examines firm behaviour	Explains growth in human capital *H* endogenously	Emphasis on growth not stock of *H*	Allows non-convergence
Solow's neoclassical model	technological change	No	No	No (no *H*)	n/a	No
MRW's augmented Solow model	level of technology	No	No	No	Yes	Yes
The AK model	productivity	Yes	No	No (no *H*)	n/a	Yes
The Lucas model	No *A*	n/a	No	Yes	No	Yes
Product variety model	No *A*	n/a	Yes	No (no *H*)	n/a	Yes
Nelson and Phelps	index of technology	Yes	No	No	No	Yes
Schumpeterian with creative destruction	productivity	Yes	Yes	No (no *H*)	n/a	Yes
Schumpeterian with two levels of human capital	productivity	Yes	Yes	No	No	Yes

What is it in the models that have been discussed that education policymakers might find most interesting or useful? One thing that stands out is a growing emphasis on human capital. To illustrate, whilst Solow's original neoclassical model did not make explicit reference to human capital, in the last of the three Schumpeterian models the level of education of the adult population is a central determinant of the mix of imitation and innovation in the economy and also the rate of income growth. The role of the education policymaker is thus central in the development process. The models include ideas on how to improve the stock of human capital that would sound familiar to informed education policymakers. Whilst the term imitation may carry a somewhat pejorative meaning implying a lack of creativity and an insufficient grasp of different contexts, carefully considered imitation is undoubtedly an element of industrial development, institutional development, the development of education systems with the right structures and incentives and indeed learning in the classroom. Effective imitation requires careful studying of how successful societies and systems produce goods and services, including education. For this, communication and information-sharing technologies such as the internet have a vital role to play. In fact, as suggested by the 'AK model' examined in section 2.4 above, the availability of technologies such as the internet can in itself influence human development. Potentially, countries with the lowest levels of development can make the easiest gains because these countries can learn from the mistakes of more developed countries. Put differently, their scope for imitating their way out of under-development is large. As Lin and Chang (2012: 484) put it, 'poor countries should be able to take advantage of their backwardness, by importing modern technology and institutions developed elsewhere'.

3 MEASURING HUMAN CAPITAL AND COUNTRY DEVELOPMENT

3.1 Introduction

Both the education and growth variables of the models that are the focus of the dissertation are associated with lively debates around how best to measure the two phenomena. This section examines these debates, which occur on both a more technical level and a more fundamental level. On a technical level, analysts have looked at how educational attainment data can be made internationally comparable when education systems are so country-specific. On the growth side, there are important questions around how to make per capita income comparable over time when, due to technological change, new goods and services emerge which are difficult to compare to their predecessors. In making price adjustments, can fixed line telephones be equated to cellular telephones, for example? More fundamental questions include whether the duration of one's past education means much when what one has learnt should perhaps be the key concern, or whether one can examine current income without considering how the generation of this income impacts on the income of future generations in a context of ecological change.

Section 3.2 focuses on the kinds of measures economists have developed to describe the amount of education received by adults in the population. As will be seen, this work, which has been tackled by a variety of economists, has not been easy, largely due to data problems.

Section 3.3 looks at measures of actual learning achieved by the adult population or pupils at a specific level of the education process. This is an area that is relatively new and has been made possible due to the increasing availability of internationally comparable and sample-based test data. Analysis of existing techniques aimed at making test results from different testing systems comparable leads to the proposal of a new approach that would make a larger variety of testing systems linkable.

Section 3.4 turns to growth or country development. Technical questions relating to the suitability of typically utilised measures of per capita income are explored. Alternatives to the typical measures, but also the data difficulties associated with these alternatives, are discussed. Two interesting alternatives discussed are subjective measures of wellness and the human capabilities measures promoted by Amartya Sen.

Finally, section 3.5 looks beyond measures of the current and past welfare of countries and individuals and brings in the matter of future sustainability. This is an area of analysis which has grown very rapidly as evidence on the relationship between economies and the ecology has strengthened, providing insights into matters such as climate change that were not available, or not widely acknowledged, as recently as the 1990s. The debate around how economists can properly model ecological impacts and the feedback of these impacts on the economy is one of the most dynamic and interesting debates within economics currently. The implications of new evidence on ecological patterns for the education planner are discussed.

3.2 Indicators of the quantity of education

Human capital is clearly a very multi-faceted concept. It embraces health, individual and collective knowledge, physical and mental skills acquired in life, mental and physical qualities individuals are born with, and cultural values that contribute towards economic progress. In the growth literature, what matters are those aspects of human capital which are changeable and which one suspects contribute towards improvements in the more economic dimensions of human welfare. Gary Becker (1993: 392), Nobel economics laureate for his work in the role of human behaviour in economic models, makes the following points about understanding human capital in terms of its returns, but also about the need to view human capital broadly:

> Human capital analysis starts with the assumption that individuals decide on their education, training, medical care, and other additions to knowledge and health by weighing the benefits and costs. Benefits include cultural and other nonmonetary gains along with improvement in earnings and occupations, whereas costs usually depend mainly on the forgone value of the time spent on these investments. The concept of human capital also covers accumulated work and other habits, even including harmful addictions such as smoking and drug use. Human capital in the form of good work habits or addictions to heavy drinking has major positive or negative effects on productivity in both market and nonmarket sectors.

Measures of the amount of education adults have received, in terms of years of education or levels of education begun or completed, have featured prominently in growth modelling from at least the work of Mankiw, Romer and Weil (1992) to more recent work such as that by Vandenbussche, Aghion and Meghir (2006). However, in more recent years there has been a shift towards the indicators of schooling quality discussed in section 3.3 below. Measures of the quantity of education received clearly have the advantage that they use data that have been relatively available for many decades for many countries and that they can deal with different levels of education, such as the completion of primary schooling and the completion of tertiary education.

Regular within-country reporting on the amount of schooling occurring has existed for some time. Benavot and Riddle (1988) were able to find enrolment data for 126 countries or colonies within the 1870 to 1940 range. However, these data suffer from serious cross-country comparability problems. Concerted efforts to ensure that countries generated internationally standardised data on the quantity of education only began after the Second World War and were led by UNESCO.

There are two indicators of the quantity of education attained by adults that have received attention amongst economists. One is average years of formal education attained by adults. Though this is often not made clear, what is actually meant by this indicator is non-repeated years of formal education received. Thus years spent repeating grades at school due to poor performance are not counted and nor are years spent repeating the same educational level through choice, something that can occur when for instance someone obtains two first degrees at the tertiary level. The indicator is thus based on the question of what the highest level of education attained is of the individual. The second indicator is the proportion of adults having completed (or even commenced) specific levels of education, such as primary schooling and tertiary education. The preferred data source for calculating both of these indicators is household surveys. However, as will be seen below, enrolment data have often been used to impute gaps in the household data.

In a few instances, enrolment data on their own have been used by economists as a proxy for levels of education attained amongst adults. This is the case in the work of Mankiw, Romer and Weil (1992), discussed in section 4.2, where historical participation in secondary schooling is considered a sufficient indicator of the level of education of the workforce in later years.

For the second of the two indicators, the International Standard Classification of Education, or ISCED, maintained by UNESCO, is important. This classification system, first published in 1997 and then reissued in 2011 with some changes (UNESCO, 2011b), defines eight education levels, from early childhood development to doctoral studies, eight education fields, examples being ‘general programmes’ and ‘social sciences, business and law’, as well as two ‘orientations’, general and vocational education.

UNESCO’s Institute for Statistics, UIS, has focussed mainly on the collection, from ministries of education, of enrolment aggregates broken down by dimensions such as educational level and gender, not on household-derived educational attainment data. The UIS online data query system[9] has annual country-level enrolment aggregates going back to 1970. These data are described in more detail in Part II of the dissertation. Enrolment data were also collected by UNESCO before 1970, but those data are not easily accessible and it is not clear why they are not included in the online query system. The literature on the calculation of country-level attainment values discussed below points to the existence of historical enrolment aggregates within UNESCO going back to around 1950. This information seems to have been published in past issues of UNESCO’s statistical yearbook. This publication does not exist any more, clearly because UNESCO has decided to rather disseminate its statistics through the UIS querying facility. The statistical yearbooks, titled in the earliest years ‘Basic facts and figures’, are registered in UNESCO’s online library catalogue, but unlike many other UNESCO documents, are not available as downloadable files. It is clearly not easy to access pre-1970 UNESCO enrolment statistics, though such historical data would have a number of uses.

Previously, UNESCO did focus a little on gathering and to some extent standardising household-derived aggregates on educational attainment amongst adults. Lutz, Goujon, Samir and Sanderson (2007: 201) used attainment statistics for adults drawn from the discontinued UNESCO statistical yearbooks. They explain that these statistics were derived from national surveys and that UNESCO made adjustments to improve international comparability. No educational attainment statistics are available on the UIS web facility. What is available online is the 1993 UN statistical yearbook, the most recent in this series to have special supplementary tables on educational attainment. For instance, this publication includes for South Africa 1985 statistics on adult educational attainment, broken down by single years of education attained and according to ten-year age cohorts. However, the statistics are of limited use as they exclude what were the four independent ‘homelands’ at the time (United Nations, 1995: 803).

An important source of household data that includes educational attainment variables is the Demographic and Health Survey (DHS) datasets. By 2012, this internationally standardised initiative had covered 84 developing countries through 366 separate surveys, beginning in 1985[10]. The microdata are downloadable in the case of around 90% of the surveys. It seems that in the remaining 10% of surveys countries decided

[9] Online data-querying facility of UNESCO: UIS, consulted September 2012.
[10] http://www.measuredhs.com/data/available-datasets.cfm. Accessed November 2012.

not to make the data public. Only developing countries are covered by the DHS, which is largely funded by the United States Agency for International Development (USAID). The DHS is clearly an important data source for cross-country comparisons if the focus is developing countries. South Africa conducted the survey in 1998 and 2003, though only the 1998 data are downloadable.

The Labour Force Survey (LFS) is another household survey initiative that aims for international comparability. Here the International Labour Organization (ILO) is responsible at the global level for standardisation and pursues this largely through promoting standard definitions of, for instance, unemployment and industrial sectors (Browne and Alstrup, 2006). It seems to be standard practice for LFSs to include a question on the highest level of education attained by each household member. Whilst the ILO archives on its website official LFS reports produced by countries, LFS microdata are not made available through the ILO. Statistics South Africa has conducted LFSs since at least 2000.

One source of household data that is emerging as an increasingly valuable source is the IPUMS[11] International online querying facility hosted by the Minnesota Population Center at the University of Minnesota. This facility is considered by Lutz, Goujon, Samir and Sanderson (2007) as having considerable potential for improving internationally comparable measures of the quantity of education. The IPUMS website[12] revealed normalised census data from 211 censuses from 68 countries. The data are provided to researchers free of charge following an application process.

Psacharopoulos and Arriagada (1986) offer what seems to be the earliest attempt amongst economists to compile a comprehensive set of educational attainment values that could be used for growth modelling. They focus on both average years of schooling and the proportion of the labour force completing specific levels of education. Using a variety of published reports emerging from national censuses and Labour Force Survey data, they managed to piece together values for at least one year, for a period beginning in 1970, for 99 countries. For 33 countries values for more than one year, mostly two years, were obtained. Their values are included in their article. One important input required for their statistics is country-specific information on the structure of the education system so that, for instance, they are able to translate a category such as 'incomplete secondary schooling attained' into years of education. In this specific example, they would assume that the respondent had completed half of the years normally associated with the secondary schooling cycle.

Kyriacou (1991) uses gross enrolment ratios calculated by UNESCO, and published in UNESCO's yearbooks, to expand on the Psacharopoulos and Arriagada (1986) dataset. Gross enrolment ratios are commonly calculated ratios where total enrolment in a level of the education system, for instance primary schooling, is the numerator and the population corresponding to that education level, assuming no grade repetition, is the denominator. Kyriacou (1991) first runs a regression with the following form:

$$H_t = B_0 + B_1 P_{t-15} + B_2 S_{t-5} + B_3 T_{t-5} \qquad (25)$$

[11] Integrated Public Use Microdata Series.
[12] https://international.ipums.org/international/index.shtml (accessed November 2012).

H_t is the average years of education in year *t* found by Psacharopoulos and Arriagada (1986). *P, S* and *T* are the gross enrolment ratios for primary, secondary and tertiary education either 5 or 15 years prior to year *t*. Kyriacou (1991) has sufficient data to include 42 countries in his regression, which produces an R^2 value of 0.82, and finds the value of the coefficients *B*. He then uses these coefficients to impute *H* for a larger number of countries and years. The end product, included in Kyriacou's paper, is a table of 117 countries, most of which have four values for *H* corresponding to different years in the period 1965 to 1985. One weakness with Kyriacou's (1991) approach is that there is no clear justification for the lags used in equation (25), for instance the lag of five years between tertiary enrolments and average years of education of adults.

Barro and Lee (2001) constitutes an update of their earlier work released in 1993 and 1996. Barro and Lee's datasets have been widely used by analysts. There are downloadable Excel files with results linked to the 2001 article. Both average years of schooling and the proportion of adults according to seven levels of attainment are provided. The seven levels include incomplete attainment, as in incomplete primary schooling. Separate statistics are provided for adults aged 15 and above and adults aged 25 and above. Values are given for 111 countries and for every five years in the period 1960 to 2000. 2000 values are future projections. Like Kyriacou (1991), Barro and Lee (2001) use UNESCO enrolment statistics to impute educational attainment data, but they use an alternative approach, the perpetual inventory approach, which is often used for physical capital. In this approach, additions to and subtractions from a stock in every time period are used to estimate the evolving total of the stock. Examining how the tertiary level is dealt with illustrates the general method. In equation (26) below, the number of adults with tertiary education (education level 3) in period *t* is the corresponding number from an earlier period, in this case *t*-5, adjusted downwards using the contents of the first brackets on the right, which capture demographic trends, and then adjusted upwards using the second term on the right, where *G* is the gross enrolment ratio for the tertiary level and the second term as a whole represents the number of tertiary graduates moving into the adult population[13]. The first subscript for each *P* refers to an age range. One data source used by Barro and Lee (2001) but not Kyriacou (1991) is United Nations population by age data.

$$H_{3,t} = H_{3,t-5}\left(1 - \frac{P_{15\to19,t} + P_{15\to65,t-5} - P_{15\to65,t}}{P_{15\to65,t-5}}\right) + P_{20\to24,t}G_{3,t} \qquad (26)$$

Around half of the final Barro and Lee (2001) data points required imputation using enrolment ratios according to the above formula. Barro and Lee (2001: 545) acknowledge that one drawback with their approach is that they assume, for instance, that if 10% of adults have tertiary education, then 10% of those dying also have tertiary education, which implicitly amounts to the assumption that educational attainment is the same at all ages. Clearly this is a problematic element of the methodology. As will be seen, subsequent analysts have strongly criticised the method

[13] More precisely, the tertiary gross enrolment G is adjusted downwards to deal with grade repeaters. This is captured in equations (4) and (5) in Barro and Lee (2001: 544-5). This adjusted ratio is considered preferable to the more or less equivalent net enrolment ratio, which is published by UNESCO. The latter statistic, they argue, is plagued by too many irregularities.

but also the usability of the Barro and Lee (2001) data, a criticism worth noting given how widely the Barro and Lee data have been used for growth regressions.

Cohen and Soto (2007a) appear to offer an improved set of figures relative to Barro and Lee (2001). Their figures cover 95 countries and the period 1960 to 2010, with values provided for every interval of ten years. The 2007 article comes with a separate data file. Average years of schooling and attainment by seven levels of education are provided, for the population aged 15 and above and the population aged 25 and above. Part of the strength of the Cohen and Soto dataset lies in the fact that they break the original household data down by age groups and thus avoid the Barro and Lee (2001) assumption that educational attainment is the same at all ages. Moreover, Cohen and Soto (2007a) make use of a much wider range of data, including new OECD household data, covering OECD members and 15 non-members, and a series of books by Brian Mitchell published in the 1990s where evidence on enrolments going back to 1750 are gathered fairly exhaustively. To illustrate, Cohen and Soto (2007a: 55) use intake ratios for primary schooling from the 1920s to estimate the attainment of people aged 60 to 64 in 1980.

Cohen and Soto (2007a), in arguing against the usability of the Barro and Lee (2001) data, point to a few glaring anomalies in the latter, for instance that in Canada the percentage of the population aged 25 and above with tertiary education moved from 6.9 to 8.2 to 7.1 in the years 1975 to 1985. This statistic is unlikely to change to such a large degree in such a short space of time. Moreover, France and Bolivia display the same average years of schooling for those aged 25 and over in 1960, of 5.4 years.

Krueger and Lindahl (2001: 1116) turn to the World Values Survey (WVS) dataset in order to verify the accuracy of the Barro and Lee data. At the time of their analysis, the WVS dataset was far more limited for this purpose than it subsequently became and it was consequently difficult for Krueger and Lindahl (2001) to draw firm conclusions in this regard. The WVS data, described further in section 3.4, has a downloadable file of all data collected in the 1981 to 2008 period. The data in this file were analysed in order to gain a sense of recent enhancements. What is clear is that the use of a question explicitly asking the highest level of education attained, as opposed to an earlier question that asked just the age up to which the respondent schooled, has become more common. Specifically, the 1981 to 2008 data include 57 countries where the educational attainment question has been asked.

Krueger and Lindahl (2001) use a simple approach to determine which of any two educational attainment datasets are more reliable. They regress, for instance, the average years of schooling values in the 1993 version of the Barro and Lee data on the corresponding values in Kyriacou (1991), and then perform the reverse analysis, and then consider the resultant slope coefficients their 'reliability ratios'. For instance, the slope coefficient on the Kyriacou (1991) values when the Barro and Lee data is regressed on the Kyriacou data would be the reliability ratio of the Kyriacou data, and vice versa. The underlying assumption is that if two variables are supposed to measure the same thing, then if one variable displays greater variance than the other, then this is the least reliable variable. The assumption is captured in the following two equations. The estimated slope coefficient in the first equation, $\hat{B}_1$, is always equal to the reliability ratio r_X for variable *X*.

$$Y_i = \hat{B}_0 + \hat{B}_1 X_i + u_i \quad (27)$$

$$r_X = \frac{\text{cov}(X, Y)}{\text{var}(X)} \quad (28)$$

Below, the Krueger and Lindahl (2001: 1116) reliability ratio values are reproduced, together with a few new ones comparing the more recent 2001 Barro and Lee dataset to the Cohen and Soto (2007b) data, and the 2001 Barro and Lee data to the downloaded World Values Survey data (World Values Survey Association, 2009). All the new comparisons use data covering the population aged 15 and above. The proportion of the population that had completed secondary and post-secondary education were selected for the WVS comparison, partly because using WVS to derive average years of schooling would have meant obtaining institutional information from all the countries concerned relating to the duration of specific levels of education. The Cohen and Soto data emerge as more reliable that the Barro-Lee data when one is dealing with completion of secondary and tertiary education. The WVS tertiary education statistic stands out as being particularly unreliable when compared to the corresponding Cohen-Soto statistic. When the WVS data were examined more closely, it was found that there were clear anomalies. A few developing countries, such as Indonesia and Dominican Republic, have unrealistically high attainment of tertiary education values in the WVS data of over 60%. In the Cohon-Soto dataset, the highest figures for the two countries are 6% and 18%. Some developed countries also have clearly incorrect values for the same statistic in the WVS dataset, for instance Australia's 60% against 25% in the Cohen-Soto data.

Table 2: Reliability of measures of educational quantity

	Reliability ratios (standard errors in brakets)	
Krueger and Lindahl (2001: 1116)		
	B&L93	*Kyriacou*
Average years 1985	.773 (.055)	.966 (.069)
Δ Average years 1965-85	.577 (.199)	.195 (.067)
	B&L93	*WVS*
Average years 1990	.903 (.115)	.727 (.093)
New calculations		
	B&L2001	*C&S2007*
Average years 1990 (n=84)	1.052 (.039)	.856 (.031)
Δ Average years 1960-2000 (n=83)	.503 (.078)	.669 (.104)
Secondary 1990 (n=84)	.864 (.051)	.897 (.053)
Δ Secondary 1960-2000 (n=83)	.403 (.067)	.763 (.127)
Tertiary 1990 (n=84)	.707 (.039)	1.134 (.062)
Δ Tertiary 1960-2000 (n=83)	.597 (.049)	1.077 (.089)
	WVS	*C&S2007*
Secondary 1990 (n=57)	.473 (.179)	.238 (.090)
Tertiary 1990 (n=57)	.106 (.051)	.694 (.331)

Sources (for the new calculations): Barro and Lee, 2000 (though the dataset is dated 2000, the journal article is from 2001); Cohen and Soto, 2007b; World Values Survey Association, 2009;

Lutz, Cuaresma and Sanderson (2008) provide a high-level summary of an ambitious effort not very different from the Cohen-Soto one, but apparently more labour-intensive. The outcome of the work is known as the IIASA-VID[14] dataset, which

[14] The two acronyms stand for International Institute for Applied Systems Analysis and the Vienna Institute of Demography.

covers 120 countries in the 1970 to 2000 period, with values corresponding to four levels of education for every five years. One big difference is that the IIASA-VID data breaks down the statistics by five-year age cohorts. The data are downloadable[15]. What is also downloadable is a related set of projections for the years 2005 to 2050. These projections follow four scenarios: a constant enrolment *numbers* scenario, a constant enrolment *ratios* scenario, a global education trend scenario, and a fast-track scenario (Samir, Barakat, Goujon *et al,* 2010). The first two scenarios are conservative and are different forms of a no-change scenario, whilst the fourth scenario represents the greatest change.

Details on how the 1970 to 2000 dataset was derived can be found in Lutz, Goujon, Samir and Sanderson (2007), which indicates that average years of schooling was calculated as part of the project, although this statistic is not included in the downloadable dataset. Conceptually, the approach followed for the attainment by educational level statistics is a relatively simple one. The year 2000 was used as the empirical base. For this year they drew from a variety of household datasets in producing their enrolment attainment statistics. Then UN population data were used to project the education information backwards from 2000. The general approach is described as a 'demographic multi-state method' (Lutz, Goujon, Samir and Sanderson, 2007: 198). The following formula explains the basic operation, for which there are a number of exceptions. Essentially a segment of the population of age group *a-5,* with highest level of education attained *e,* at point in time *t*-5, of a particular sex *s*, is calculated by taking the corresponding group five years later and dividing this by a survival ratio *s*.

$$P_{a-5,e,t-5,s} = \frac{P_{a,e,t,s}}{s_{a-5,e,t-5,s}} \qquad (29)$$

The data sources used to establish the breakdown of educational attainment by age in 2000 (or two years either way, which was adopted as a leeway) included national census datasets, DHS datasets, Eurostat datasets (Eurostat is the statistical body of the European Union) and LFS datasets. The DHS data proved to be the most difficult to use in terms of aligning education attainment data to ISCED levels.

A greater hurdle was to establish what education-specific mortality rates to use for developing countries, where these statistics were not readily available. The solution adopted was to analyse household datasets for different points in time for eight countries, seven of which were developing countries, and in this way to establish patterns of education-specific mortality. This analysis then informed what mortality rates to use across those countries without easily available statistics. The basic pattern found across several countries was that life expectancy at age 15 became a year longer when no schooling became primary schooling completed. Life expectancy rose by a further two years if secondary schooling had been completed and by another two years if tertiary education had been completed. What was clear was that using the same mortality rates for differently educated people was not good modelling practice. The education-specific life expectancy values were used to generate education-specific mortality rates at the different five-year age groups, but in such a way that the original age-specific mortality rates in the UN population data were respected.

[15] http://webarchive.iiasa.ac.at/Research/POP/edu07/ (accessed November 2012).

Lutz *et al* (2007) originally considered modelling migration but eventually abandoned this due to data constraints and the fact that it was clearly not possible to generalise across countries. How education influences in- and out-migration appears to be very country-specific. However, it was also concluded that not modelling migration explicitly did not affect the results significantly.

Comparability problems over time within countries are reduced because in any year the classifications will correspond to what applied in the base year, ideally 2000. The typical concern around how to deal with structural changes in the education system over time did thus not apply.

What are the key advantages of the IIASA-VID data relative to the other education datasets that have been discussed above? One is that the IIASA-VID statistics are disaggregated by age bins, so it is possible to examine separately the effects on growth of younger and older segments of the population, according to educational attainment. This type of growth modelling is in fact done by Lutz, Cuaresma and Sanderson (2008) and is discussed in depth in Part II of the dissertation, where it is relevant because the modelling allows for the impact of different levels of the education system to be assessed. The Cohen and Soto (2007b) data could be made available with disaggregation by age, because Cohen and Soto model the data using age bins. However, their statistics were not presented in this way, but rather for adults in general. Yet one fundamental drawback of the Cohen-Soto data relative to the IIASA-VID data is that the former implicitly assumes that the mortality rate is the same, regardless of the educational background of the population.

3.3 Indicators of schooling quality

Indicators of educational attainment can at best be proxies for what economists are really interested in in the growth process as far as human capital is concerned: the knowledge and skills of the labour force or, probably more accurately, those directly or indirectly involved in the production of goods and services in the economy, plus those who might have been involved had their human capital endowments been better. The latter definition, which essentially amounts to all adults, acknowledges the importance of, for instance, women not directly engaged in the labour market and the wastage of human capital implied by unemployment.

Defining what skills and knowledge one should ideally measure in order to understand linkages between education and growth is a debatable matter. Heyneman (2004a) provides a useful outline of the debate. There are important skills trade-offs that must be considered. The trade-off between the promotion of critical thinking and the fine-tuning of more mechanistic thinking in the education system is often mentioned, for instance in contrasting the resultant human capital of Western Europe, on the one hand, and ex-communist Eastern Europe or East Asia, on the other. The latter tend to surpass Western Europe in mathematical skills, whilst Western Europe tends to perform better when it comes to complex problem solving involving critical thinking. Different stages of country development may require different emphases in the education system. Another trade-off is one between the number and depth of skills (or knowledge) possessed by adults. To illustrate, Heyneman (2004a) contrasts the emphasis in the French mathematics school curriculum on a wide range of topics to the Algerian emphasis on fewer topics, presumably covered in greater depth. Any attempt to gauge the skills levels of the population must consider these issues.

Heyneman (2004a) also discusses the impact of the rise of international trade in education services, in particular at the tertiary level, on notions of measuring human capital. In a context where, for instance, around a fifth of the enrolment in the top universities of the United States is not from the United States, it becomes problematic to use measures derived from university students enrolled in a country to gauge the quality of human capital amongst the adult population of that country, not only with respect to the United States, but even countries from which students originate.

One clearly non-ideal measure of the human capital of adults is the adult literacy rate found in the UIS's online data querying facility. The data source for adult literacy rates has traditionally been simple questions included in household surveys where subjectivity is a major problem. For example, the 2011 national census in South Africa has the question: 'Does ... have difficulty in reading (e.g. newspapers, magazines, religious books, etc.) in any language?'[16]. As noted by Gustafsson, Van der Berg, Shepherd and Burger (2010), in a paper that attempts to gauge the costs of illiteracy to South Africa, not only are there unavoidable across country comparability problems when self-reported literacy data are used, there are also within-country comparability problems over time. With respect to the latter, Gustafsson *et al* (2010) point to changing relationships between highest school grade attained and self-reported literacy which suggest that over time respondents, at least in South Africa, become more demanding with respect to the latter variable, in other words less inclined to claim that they are literate. A few economists, notably Romer (1989), have used UIS literacy rates on the right-hand side of growth regressions. Romer (1989), unsurprisingly given the comparability problems in the data, finds adult literacy to have no statistically significant conditional correlation with growth.

By far the most promising efforts to resolve information gaps around the human capital of countries are those of the OECD focussing on testing random samples of adults through visits to households. This work has been limited to OECD countries, yet its cost-effectiveness makes it likely that it will increasingly occur in developing countries too.

The OECD's work has occurred within three consecutive testing programmes which reflect a series of methodological lessons learnt and adjustments. The first programme, the International Adult Literacy Survey, or IALS, involved data collection from 21 countries during the 1994 to 1998 period. Each test event was conducted by a fieldworker and involved about an hour-long test, taken by adults aged 16 to 65, which covered both very basic literacy skills, in order to allow for differentiation amongst those with low literacy levels, and more advanced skills in three areas: comprehension of short narrative texts, comprehension of other information sources such as maps, tables and graphs, and the solving of number problems people typically encounter in their lives[17]. IALS was succeeded by the Adult Literacy and Lifeskills (ALL) survey, which involved data collection in 2003 and 2006 in 12 countries, mostly countries that had already participated in IALS[18]. A key enhancement here was a stronger emphasis on skills needed to use modern information and communication technologies (ICTs). Finally, the OECD introduced

[16] The census questionnaire is available at http://www.statssa.gov.za.
[17] The official IALS report is OECD (2000).
[18] The ALL report for the 2003 wave of testing is OECD (2005). A final report for the 2006 collections was not found.

the Programme for the International Assessment of Adult Competencies (PIAAC), which involves an initial wave of data collection across 28 countries in 2011 and is expected to produce a final report in 2013. PIAAC involved further strengthening of the testing of ICT skills as well as the introduction of a preferred option of testing where the test-taker would enter responses on a laptop computer, as opposed to on a paper form. Thorn (2009), in an evaluation of the three programmes, points out a few concrete examples of the value they have added to policy formulation, in particular within the area of national adult literacy programmes.

Coulombe, Tremblay and Marchand (2004) used the 1994 IALS data from 14 countries to construct a historical panel dataset for the period 1960 to 1995 of the human capital of 17 to 25 year olds, in other words new labour market entrants, using the assumption that the literacy levels of older adults in 1994 reflected the literacy levels of younger adults in earlier years. The study included growth regressions where the test-based literacy data were significantly correlated with growth. This was seen to underscore the importance of having reliable human capital data in one's growth models. The data also allowed for more detailed findings, for instance that average literacy is more important for growth than top-end literacy, or achievement amongst the best test-takers. Interestingly, it was found that female literacy displayed larger coefficients in the growth models than male literacy.

Whilst by 2012 there appeared to be no concrete and decisive moves towards internationally standardised adult literacy testing in non-OECD countries, what had grown phenomenally in all countries was participation in standardised testing programmes aimed at the primary or secondary schooling levels. This is understandable given the centrality of schooling in the policy agendas of developing countries and the relatively marginalised state of adult literacy programmes in many of these countries.

Hanushek and Woessman (2009) have made extensive use of the emerging school-level data. Their work on creating comparability across the data of different testing programmes is discussed here, whilst their use of the normalised data in growth regressions is discussed in section 4.4.

Hanushek and Woessman (2009: A14) combine country average scores across disparate international testing programmes, around 800 country scores pertaining to 77 countries, into one set of comparable scores, one per country, that can be used within a cross-country growth model[19]. Just over half of the 77 countries for which comparable scores are obtained are high income countries. The original data were obtained from the NAEP[20] programme of the United States, the OECD's PISA[21] programme, and the IEA's TIMSS, PIRLS[22] and initial programmes, where the initial programmes were essentially the precursors to TIMSS and PIRLS. The original country scores are all from the period 1964 to 2003, with the mean year of collection being 1994 (actual years of collection were weighted by the number of data points per

[19] Hanushek and Woessman provide a slightly different version of the same calculations relating to the 77 countries in OECD (2010a: 36).
[20] National Assessment of Educational Progress.
[21] Programme for International Student Assessment.
[22] Trends in International Mathematics and Science Study and Progress in International Reading Literacy Study, both programmes of the International Association for the Evaluation of Educational Achievement.

year). Each score is specific to a subject, a grade (or age level) and a test year. There were test scores for three subjects, mathematics, science and reading, and for three institutional levels, primary, lower secondary and upper secondary. The original data are from 32 different tests, where a test is specific to a year, an age (or grade level) and a subject. For instance, the 2003 TIMSS test in mathematics for grade 8 would be a 'test'.

A key phenomenon that greatly facilitates the analysis is the fact that correlations between scores from different programmes with a different testing focus for the same subject and grade (or age level) are in fact high where countries participated in more than one programme. In particular, the correlations between PISA and TIMSS country averages for the same subject are found to be in the range 0.86 to 0.97 (Hanushek and Woessman, 2009: A3) despite the fact that TIMSS is more oriented towards existing curricula whilst PISA is more oriented towards the application of knowledge to the real world.

Hanushek and Woessman (2009: A4) argue that one factor that complicates the analysis is the fact that the Item Response Theory (IRT) approach used in the later years causes incomparability across years. The technical documentation from the programmes themselves argues convincingly that *within* programmes there is not always a comparability problem as long as one is dealing with the same subject and grade (or age level). For instance, Mullis *et al* (2008: 402) explain that in the TIMSS mathematics runs after 1995 scaling occurred in such a way that scores were made comparable to the 1995 baseline scores. Thus, for instance, the mean of 500 that is referred to when TIMSS 2007 country results for grade 4 mathematics are presented (Mullis *et al*, 2008: 34) is the mean for the 1995 run, not for the 2007 run. This is what allows for comparisons across years in the official TIMSS reports. Similarly, in PISA, reading scores from the 2000, 2003, 2006 and 2009 runs are comparable, though in the case of mathematics and science results for 2000 are not comparable to those in later years, and in the case of science even for the later years special country averages other than the official ones must be used for the comparison to be valid (OECD, 2009: 157; OECD, 2010b: 21, 186). There is thus comparability across years within the same programme sometimes, but not always.

The method used by Hanushek and Woessman (2009) to arrive at normalised country scores is strongly dependent on the fact that the United States participated in every test considered in the analysis. In a nutshell, the method involves comparing each country's score to that of the United States in the same test, and then using the United States NAEP results, which are available for virtually every year, age level and subject in question, to bring about adjustments that will make all scores comparable over time. NAEP results, according to the analysts, provide the only set of results that are truly comparable over the entire time period. NAEP tests are only run in the US.

A key end product of the normalisation process published by Hanushek and Woessman (2009: A15), and summarised here with some simplifications, is a general and comparable measure of cognitive ability per country, where this measure is an average across two subjects, mathematics and science, across all three institutional levels and spanning all the years for which the country had data. The main focus is thus not on changes in educational quality over time but rather differences in the levels of educational quality across countries in implicitly the same time period. In fact, both data availability and caveats relating to the calculation make trends across time difficult to discern.

The aim of the approach is to transform all country scores to the scale followed by PISA in 2000. Each transformed country score *T* can be described as a linear function of the original country score *O*.

$$T = \alpha + \beta O \tag{30}$$

The slope coefficient β in the case of a specific test can be said to be obtained according to equation (31) below. Here the original notation is used, but with the terms reorganised to display the linearity of the function.

$$\beta_{a,s,t} = \frac{SD^{OSG}_{s,PISA,2000}}{SD^{OSG}_{a,s,t}} \tag{31}$$

The original country score *O*, for institutional level *a*, subject *s* and year *t* is converted to the PISA 2000 scale through multiplication by the ratio of the standard deviation across country scores in PISA in 2000 in subject *s* to the standard deviation across country scores with respect to *O*. But this ratio is based only on a subset of countries belonging to the 'OECD standardization group', or OSG. This is a group of 13 countries which displayed consistent participation across many tests.

The intercept α for the test covering institutional level *a*, subject *s* and year *t* is obtained as follows:

$$\alpha_{a,s,t} = O^{US}_{s,PISA} - O^{US}_{a,s,t}\left(\frac{SD^{OSG}_{s,PISA}}{SD^{OSG}_{a,s,t}}\right) + \left(NAEP^{US}_{a,s,t} - NAEP^{US}_{a,s,1999}\right)\frac{SD^{US,PISA}_{s}}{SD^{US,NAEP}_{a,s}} \tag{32}$$

The first term on the right-hand side is the US country score in PISA in 2000 in the subject in question, *s*. From this is subtracted the original US country score for the test in question converted to the PISA 2000 scale using the ratio referred to previously. The third term is the change over time in the US NAEP mean converted to the 2000 PISA scale using a ratio of standard deviations of scores across United States pupils. If the test in question is a 2000 PISA test then α reverts to zero (the difference between the two NAEP means would be zero as there was no NAEP test in 2000 and in its place the 1999 NAEP results would be considered).

A specific limitation brought about by the requirement that the United States be a participant in every test is that the regional programmes SACMEQ[23] and PASEC[24] in Africa and SERCE[25] in Latin America cannot be considered. The SACMEQ and SERCE regional programmes include 19 developing countries which do not participate in programmes where the US participates. PASEC includes a further 11 such countries, but PASEC statistics are not used in the data normalisation process described below because there are no PASEC countries which also participate in some

[23] Southern and Eastern African Consortium for Monitoring Educational Quality.
[24] Programme for the Analysis of Educational Systems of CONFEMEN, where CONFEMEN is the Conference of Education Ministers of Countries Using French in Common (from French).
[25] Second Regional Comparative Study (from Spanish).

other international programme[26]. SACMEQ and SERCE (but also PASEC) are similar in many ways to the programmes considered by Hanushek and Woessman (2009): they focus on mathematics and reading and, in the case of SACMEQ and SERCE, they have used a scale where the mean across participating countries is 500 and the standard deviation across all countries at the pupil level is 100. All three programmes are limited to the primary level, however.

An approach to obtaining normalised country scores in a manner that permits the inclusion of programmes such as SACMEQ and SERCE is the focus of the discussion that follows[27]. As in the approach of Hanushek and Woessman (2009), in the new approach described below the relationship between the original score O and the transformed score T is considered linear, implying the solution must involve finding α and β within each relationship. The assumption of a linear relationship, both in the existing and new approaches, implies an underlying assumption, namely that the shape of the distribution of country scores with respect to the same set of countries in different tests should be similar.

The viability of the approach proposed here depends on a requirement being fulfilled with respect to the pattern of bridge countries in the data used, where bridge countries are countries participating in more than one programme. The following table describes the data. Country scores from 20 tests, each represented by a year in Table 3, were used. All tests are from the period 2000 to 2009[28]. In total 742 country scores from 113 countries were used. In the discussion that follows ‘programme’ means a group of tests within which scores are comparable. PISA is one programme because in the period in question scores within a subject are comparable over time by design. Moreover, PISA mathematics and reading scores are comparable insofar as they follow the same scale of a mean of 500 and standard deviation of 100, are anchored in a specific year, and involve the same countries in any one year. Similarly, PIRLS, SACMEQ and SERCE are each counted as a programme. TIMSS grade 4 and TIMSS grade 8 are counted as two separate programmes, however, as different sets of countries participate. Around half of the TIMSS participants in 2003 and 2007 opted for testing in just one of the two grades.

[26] PASEC country scores can be found in CONFEMEN (2012).
[27] The approach is also published in the working paper Gustafsson (2012).
[28] Key datasets that became public after the analysis presented here was done are the 2011 TIMSS and PIRLS datasets.

Table 3: Summary of the test data

Test series	*Years*	*Coun-tries*	*2-point series*	*3-point series*	*4-point series*	*Total series*
PISA mathematics (age 15)	2000, 2003, 2006, 2009	72	14	10	35	59
PISA reading (age 15)	2000, 2003, 2006, 2009	72	14	11	34	59
PIRLS reading (Grade 4)	2001, 2006	47	28			28
TIMSS mathematics (Gr 4)	2003, 2007	41	22			22
TIMSS mathematics (Gr 8)	2003, 2007	61	36			36
SACMEQ mathematics (Gr 6)	2000, 2007	14	13			13
SACMEQ reading (Gr 6)	2000, 2007	14	13			13
SERCE mathematics (Gr 6)	2006	16				0
SERCE reading (Gr 6)	2006	16				0
OVERALL		113	140	21	69	230

Sources: OECD (2001, 2004a, 2004b, 2007, 2010b) and Walker (2011) (for PISA); Mullis et al (2007) (for PIRLS); Mullis et al (2004, 2008) (for TIMSS); Makuwa (2010) (for SACMEQ); UNESCO (2008) (for SERCE).
Note: In the total row there are more 2-point series than countries because some countries have more than one 2-point series. There are no overlapping series in the count of series, meaning for instance that no 3-point series would also be counted as two 2-point series.

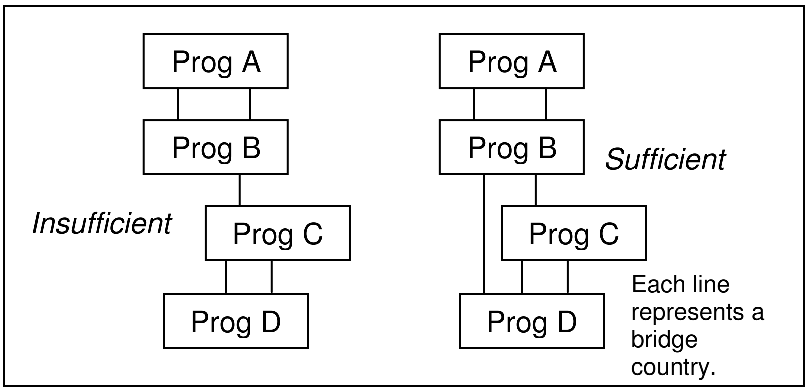
Table 4 indicates the number of bridge countries that join the six programmes to each other. For instance, two bridge countries join SACMEQ to TIMSS grade 8, namely South Africa and Botswana. South Africa also joins SACMEQ and PIRLS. Mauritius joins SACMEQ to PISA. Compared to SACMEQ, SERCE has a greater number of bridge countries joining this programme to the large international programmes. PASEC has no bridge countries that permit linking of the PASEC programme to the other programmes.

Table 4: Programmes and number of bridge countries

	PIRLS	PISA	SACMEQ	SERCE	TIMSS Gr 4	TIMSS Gr 8
PIRLS						
PISA	39					
SACMEQ	1	1				
SERCE	2	9	0			
TIMSS Gr 4	31	29	0	2		
TIMSS Gr 8	35	39	2	3	36	

The only requirement for the approach described below is that it should not be possible to divide the programmes into two sub-sets separated by just one bridge country. Any two sub-sets of programmes must always be joined by at least two bridge countries. If this condition is not met, the nonlinear programming solution explained below will not work as there would essentially be two separate optimisation processes instead of one. Figure 14 illustrates the requirement. Table 4 provides sufficient information to indicate that the requirement is met with respect to the data described in Table 3. Even if SACMEQ and TIMSS grade 8 had not been joined by any bridge countries, the requirement would still have been met, because South Africa joins SACMEQ and PIRLS and Mauritius joins SACMEQ and PISA. In fact, the bridging of SACMEQ to PISA, something which did not occur before Mauritius entered PISA, for the 2009 PISA wave, is especially valuable as it permits the anchoring of SACMEQ results at the top of the SACMEQ range (Mauritius obtained the best scores in the 2007 wave of SACMEQ).

Figure 14: Programmes and bridge countries



In the new approach, one programme must be selected to provide the standard scale for the transformed country scores. For this, PISA was selected, following Hanushek and Woessman (2009). For each country and programme, the mean of the existing country scores was calculated. For instance, one SACMEQ mean value for South Africa was calculated from the four original country scores for the two years and two subjects.

Transforming the country- and programme-specific values to comparable country scores involved solving a nonlinear programming problem whose objective function was the following:

$$\min z = f\left(\alpha_m, \beta_m, \ldots, \alpha_{m=n-1}, \beta_{m=n-1}\right) \tag{33}$$

The value z, which must be minimised, is a function of a number of pairs of intercepts, α, and slope coefficients, β, which together constitute the decision variables of the problem. The number of pairs is equal to the total number of programmes minus one, as one programme has been selected as the standard. There are n programmes in total, including the standard programme (n equals 6 in our case).

Function f can be described as follows:

$$z = \sum_{j=1}^{k} \left(\overline{D}_j W_j\right) \tag{34}$$

$$D_{ij} = \left(T_{i,m=c} - T_{i,m=d}\right)^2 \tag{35}$$

$$T_{im} = \alpha_m + \beta_m O_{im} \tag{36}$$

Value z is the sum of k values, k being the available bridges in the sense of non-zero values in Table 4. The variable k equals 13 in our case. D is a value attached to each country within a bridge. Equation (34) involves finding the mean for D across countries within each bridge (for instance the mean across 39 values within the PISA-PIRLS bridge) and multiplying this mean by a weight W (explained below). In equation (35), D for country i and bridge j is obtained by finding the squared difference between transformed country scores T for the two programmes c and d, the programmes joined by bridge j. In equation (36) the transformed country score for country i and programme m is a linear function of the original score for the country in the programme (as explained earlier, this original score can be the mean across several tests if more than one test exists).

The weight *W* attached to each of the *k* bridges is calculated in such a way that a weight of 1.0 for each country (this 1.0 is not the same as *W*) is distributed across those bridges that join the programmes in which the country participates. The sum of all the *k* values of *W* is thus equal to the number of countries, 113 in our case. The weighting system therefore follows the principle generally followed in cross-country analysis, which is to assign an equal weight to every country. Without the weighting system, in other words if every bridge carried an equal weight, countries which are found repeatedly across many programmes would implicitly tend to carry too much importance in the nonlinear programming solution. A bridge is weighted more if (1) there are more countries in the programmes being bridged, (2) if the countries in the programmes being bridged have a lower presence in other programmes and (3) if the programmes being bridged have fewer bridges serving them. The following two equations describe how *W* is calculated.

$$F_m = \sum_{i=1}^{e} \frac{1}{P_i} \tag{37}$$

$$W_j = \left(\frac{F}{B}\right)_{m=c} + \left(\frac{F}{B}\right)_{m=d} \tag{38}$$

P_i is the number of programmes that country *i* participates in. The value *F* for programme *m* is the sum of all the reciprocals of *P* for the *e* countries participating in programme *m*. Weight *W* for bridge *j* is the sum of two values, one for each of the two programmes being bridged. The programme-specific value is *F* over *B*, where *B* is the total number of bridges that serve programme *c* (or *d*).

Returning to equation (33), what the nonlinear programming solution does is adjust the coefficients of the five (n-1) linear relationships between *T* and *O* in a way that minimises the differences between the values *T* for the same country across different programmes. Because in effect $T=O$ in the case of PISA (or $\alpha=0$ and $\beta=1$), all country scores *T* gravitate towards the scale used by PISA. If the differences between the country scores of different programmes were only a matter of different programme-specific means and standard deviations with respect to the scoring process, then the difference in equation (35) would in all instances be zero and *z* in equation (33) would be reduced to zero.

The nonlinear programming problem is subject to just one constraint, which is that the transformed country scores *T* cannot be negative.

The results of the application of the new approach, in terms of programme-specific values *T* and the overall mean for each country, are provided in Appendix A[29]. The 113 overall mean values can be considered comparable country scores.

In exploring alternatives, a quadratic as opposed to a linear form was attempted for equation (36). This did not appear to come with any noticeable benefits, for instance

[29] The Excel file used for the calculations can be obtained from the author.

in terms of the performance of the country scores in the growth models described in section 4. The use of the weight *W* did not make a large difference to the results. The largest difference was in the case of Zambia, where the transformed country score would have been 265 instead of the 261 reported in Appendix A had unweighted bridges been used. Importantly, though, the goodness of fit, in terms of R^2, of the regression models discussed below was marginally better with country scores produced using weighted bridges, compared to unweighted bridges.

What are the key differences between the nonlinear programming approach described above and the approach used by Hanushek and Woessman (2009)? Two key differences stand out, both of which point to weaknesses in the nonlinear programming approach, or the cost of having more countries in one's set of normalised country scores. One is that the approach presented here is applicable at a high level in the sense that it subsumes details relating to year, subject and institutional level within the relatively untransparent decision variables of equation (33). The approach is thus not amenable to an analysis of trends over time in the way the Hanushek and Woessman approach (2009) is. A second key difference is the dependence of the approach presented here on a weighting system, the design of which is at least partly a matter of subjective judgement.

The standardised testing systems targeted at the education system and discussed above, all focus on primary and secondary schooling. What is relatively under-developed are testing systems focussing on the pre-primary and post-secondary levels. Pre-primary testing systems could play an important role in informing policy dealing with this critical level of education. For understanding linkages between education and growth, however, comparable data on how well tertiary education institutions develop skills at the top end of the human capital spectrum would be important. In Part III of the dissertation an unusual and innovative programme in Brazil that focuses on the standardised testing of university students completing their first degree will be discussed. Such a programme seems to offer valuable lessons for other countries wishing to improve within-country monitoring of tertiary education quality. It might even be a precursor to internationally standardised testing at the tertiary education level, though such testing still seems a relatively distant prospect, for both developing and developed countries. Yet, as Heyneman (2004a) argues, the defining and measuring of quality at the tertiary education level is likely to open up as an important new frontier in education research in the near future.

3.4 Economic growth indicators and their alternatives

The notion that human welfare is reflected in the level of income of people is one that has been facilitated by the emergence of market economies. The more the goods and services that people need are traded and thus have prices, the easier it becomes to measure human welfare according to a single monetary metric. It should thus come as no surprise that the earliest relatively reliable income per capita figures are those of the early industrial countries, such as Britain, for which figures from 1700 are available (see Romer (1986)). Arguably, the reliance on income as a measure of welfare in market economies has been stronger in the popular media and amongst policymakers than amongst the economists who developed the income per capita measures in the first place. Even Simon Kuznets, 1971 Nobel laureate for his empirical work on development and economic growth and a keen analyst of income per capita, was sceptical about the use of income in the absence of considerations around the quality of products, income inequalities and pollution (Kuznets, 1980). (The reason why Kuznets does not receive more attention in this dissertation, despite his prominence in the area of modelling economic development, is that Kuznets did not pay much attention to the role of human capital.)

There is a rich literature on why income per capita is a limiting measure of human welfare and on possible alternative measures. Three strands in this literature can be identified. Firstly, better ways of defining and measuring income and the related phenomena of consumption and wealth have been examined. Secondly, certain analysts, using the first strand as a point of departure, have questioned the very notion that, in particular, consumption mirrors welfare and have sought alternatives, partly using tools and concepts from the fields of psychology, sociology, political science and philosophy. Thirdly, an important strand of enquiry has emerged in recent years which again uses the first income strand as a point of departure and focuses not just on consumption that is realisable in the near future but on consumption that can be sustained over a longer term. Conceivably, it is possible to merge the three strands within a common framework for understanding welfare. However, work on this appears to be limited. Essentially the critics of the traditional income approach to welfare, at least within economics, have tended to focus either on questioning the monetary metric without paying too much attention to sustainability, or else on whether income is sustainable over time without paying too much attention to whether income is the right metric. The division of the debate into the three strands described here can be found in a recent and rather comprehensive review of the debates, compiled by a team that included two Nobel laureates, Joseph Stiglitz (2001 laureate for work on information asymmetries in markets) and Amartya Sen (1998 laureate for work on welfare economics). The review, Stiglitz, Sen and Fitoussi (2009), is henceforth referred to as SSF.

The first two strands of the discussion are dealt with in the present section, whilst the third one, on sustainability, is dealt with in section 3.5.

Beginning with the first strand, on better ways of carrying out the monetary measurement, key omissions inherent in the traditional GDP per capita measure can be grouped under nine headings: (a) international economics, (b) technological change, (c) wealth, (d) non-market production, (e) inefficiencies in government services, (f) institutional failure and defensive spending, (g) leisure time, (h) the cost of risk, (i) income inequality. Below, how taking each of these nine omissions into account represents a departure from traditional GDP per capita is discussed.

Before the discussion begins, however, it is useful to restate a few basics on how GDP per capita is calculated. The rules for ‘national accounts’ are laid down by System of National Accounts (SNA) maintained by the United Nations Statistics Division and partner organisations such as the International Monetary Fund (United Nations Statistics Division, 2009a). GDP is one of most important statistics appearing in these accounts. Three definitions of GDP can be identified depending on one’s point of departure: total economic output, total expenditure or total income. The three definitions are designed to yield the same overall GDP value. The income approach to GDP, which of the three approaches is the approach most relevant for understanding welfare, is stated as follows: ‘Income-based GDP is compensation of employees, plus taxes less subsidies on production and imports, plus gross mixed income, plus gross operating surplus'[30]. Gross mixed income and gross operating surplus are essentially earnings received by owners of unincorporated (generally smaller) and incorporated (generally larger) firms respectively. Taxes are included as roughly it is assumed that these become indirect income in the form of government services. A key question is how non-market production common in developing countries is treated. For instance, would vegetables grown by a household for its own consumption not be counted within GDP, even if similar vegetables purchased on the market by another household, using income earned, clearly would? The question is not as readily answered as one might expect. The stated intention within the SNA is clearly to count all non-market activities. In explaining the elements of the SNA, the following is said (United Nations Statistics Division, 2009a: par. 2.40):

> Given the fundamental role played by the market in modern economies, the SNA distinguishes, as an essential feature of its structure, between establishments which are market producers, producers for own final use and other non-market producers. Market establishments produce mostly goods and services for sale at prices which are economically significant. Producers for own final use produce mostly goods and services for final consumption or fixed capital formation by the owners of the enterprises in which they are produced. Other non-market establishments supply most of the goods and services they produce without charge or at prices which are not economically significant.

The vegetables for own consumption described above would be dealt with under 'producers for own final use' whilst charitable work is an example of what would fall under 'non-market producers'. 'Market producers' includes producers considered formal and informal, as long as goods are traded in a market. In the case of South Africa, Davies and Thurlow (2009: 14) find that informal market producers account for 7% of GDP. However, what is clear is that neither 'producers for own final use' nor 'non-market producers' are considered in South Africa's official GDP, despite this being the intention of the SNA. The South African practice is in line with international practice. Scandinavian countries did at one point count non-market household production within their official GDP figures, but this practice was dropped in 1950 (Ironmonger, 2001: 9).

In South Africa the calculation of GDP is a responsibility shared by Statistics South Africa, which follows the output approach, and the South African Reserve Bank, which follows the income and expenditure approaches (Statistics South Africa, 2010: 13). Collaboration between the two bodies means that the different approaches lead to the same GDP values. In its reports, Stats SA sub-divides GDP according to ten different industrial sectors. One of these sectors is 'general government services', which would include public education. The Reserve Bank, in presenting the income version of GDP, uses employee income and the income of owners of businesses as key sub-divisions. The expenditure version of GDP uses the sub-divisions household consumption, government consumption, investment expenditure and international trade adjustments (see for instance South African Reserve Bank, 2009: S104). The details behind the three different calculations of GDP are not easily available to those outside the two organisations. Little is published on the internet. In contrast, the Bureau of Economic Analysis, the body in the US responsible for drawing up the national accounts, publishes a relatively wide range of guides intended for specialists

[30] http://unstats.un.org/unsd/sna1993/glossform.asp?getitem=206.

and non-specialists on its website[31]. This is likely to reflect a wider pattern whereby countries with higher levels of public accountability and education see more technical information on the calculation of GDP being made public.

A discussion of the nine omissions referred to above follows.

(a) International economics. To deal with the problem that GDP includes, for instance, income earned by those living outside South Africa from investments in South Africa and excludes income earned by South Africans from investments elsewhere, typically one resorts to gross national income (GNI), also referred to as gross national product (GNP), a measure that adjusts for these problems so that only the income of those living in South Africa is reflected. (The SNA recognises the term GNI but not GNP.) SSF (2009: 94) prefer going a step further and using national disposable income, a term recognised within the SNA. In this indicator the value of remittances is attached to the receiver of the remittances, not their giver. Thus, for instance, the total incomes of Lesotho migrant workers in South Africa would be counted within South Africa's GNI, but South Africa's national disposable income would exclude the part of this income sent to family in Lesotho as remittances. The argument is that national disposable income would more accurately reflect the consumption and hence welfare of the migrant workers in South Africa as well as the families of these workers remaining in Lesotho. The data availability problem is that whilst GNI values per country are readily available, the same cannot be said for national disposable income. For instance, GDP and GNI figures but not national disposable income figures for many countries are available on the online data querying facility of the UN Statistics Division[32]. An analyst may be able to access sufficient national disposable income values through national sources, but this would be costly, apart from the outcome being unpredictable. As an example of this, the South African Reserve Bank, in its Quarterly Bulletin, provides values for the country's 'gross national disposable income', which in the period 2006 to 2009 have generally been around 4% lower than the corresponding GDP values. To compare, South Africa's GNI has tended to be around 3% lower than GDP in the same period. These differentials reflect the net outflows of income from South Africa, specifically through the remittances from migrant labour and profits paid to foreign investors.

(b) Technological change. Models examining economic growth by country rely on the comparability of income values across countries and over time. In order to establish comparability across countries, income is generally measured in purchasing power parity (PPP) US dollars. This measure adjusts exchange rates found in the market for the fact that the same amount of money has a different purchasing power in different countries. For instance, food and labour tends to be cheaper in developing than developed countries due to factors such as legal or logistical restrictions in the movements of goods and services across borders. At the same time, technologically advanced services such as certain medical operations can be expected to cost more in less developed countries where the necessary technology is less widespread. The estimation of the PPP relationship between different currencies occurs rather infrequently and within a research project that is costly. It was only in 1993 that a comprehensive set of PPP relationships spanning most countries was obtained (World Bank, 2008: 169). As pointed out by Coulombe, Tremblay and Marchand (2004: 11),

[31] See for instance http://www.bea.gov/national/index.htm.
[32] http://unstats.un.org/unsd/snaama/introduction.asp.

these new data were what lay behind the surge in cross-country growth modelling that occurred at the time. This PPP analysis exercise was repeated only after twelve years, in 2005. The work is largely led by the statistical bodies of the UN, though the University of Pennsylvania (the producer of many of the statistics used in section 4) was initially involved to a large degree. Income in PPP dollars is published for every year, for instance by the University of Pennsylvania. However, the further the year is from the year of the last comprehensive review of world prices, the more tenuous the comparability of the values across countries as they would become increasingly reliant on techniques such as inflation adjustments and decreasingly reliant on differences in the actual purchasing power of money relative to a basket of goods. For instance, whilst 2005 average income in PPP dollars across countries would be strongly influenced by the empirical work on prices in 2005 of actual goods, the later 2010 values would rely partly on the accuracy and comparability of nationally generated inflation indices used to make 2005 and 2010 prices comparable. The figures used for the growth models seen in section 4 are therefore less reliable than one may initially think.

What is even more elusive than the comparability of income across countries, is the comparability of income over time, even within one country. At the heart of this problem is the problem of understanding the welfare implications of technological change or, as SSF put it, of quality change. Whilst the consumer price index (CPI) for a country will allow for the comparability of income over time with respect to that income's ability to purchase non-changing goods, such as fruit and vegetables, complications arise insofar as goods and services that are new or change fundamentally are concerned. To illustrate, the computer that can be bought for 1,000 US dollars in 1990 will be very different to the computer that can be bought for the same amount in 2010, even in the absence of price inflation. Moreover, in 1980 it would have been impossible to buy a computer for USD 1,000. The question is whether the improvements to the quality of life supposedly brought about by computers and technological change generally should somehow be quantified when income or purchasing power is compared over time. Furthermore, it could be argued that certain technologies that did not exist previously, such as the cellular phone, become basic necessities over time, implying new costs for a basic quality of life that did not exist before. Can these technological considerations be incorporated into, for instance, measures of income growth over time so that these measures more accurately reflect developments in human welfare? To date, there is insufficient agreement around a generally implementable approach to doing this and even whether it is possible to introduce this level of complexity into the international accounting systems. A few notable analyses exist, however. The 1996 Boskin Commission Report of the US government estimated that annual inflation had been over-estimated by about 0.6 percentage points over about two decades because the benefits brought about by quality change had been under-estimated (United States: Social Security Administration, 1996: Section V). However, global indicators on income improvements adjusted for the quality benefits inherent in technological change are still only a remote possibility. This obviously limits the analysis into how human capital development contributes towards welfare, especially if this analysis focuses on trends over time.

(c) Wealth. Wealth, or property or capital owned, is undoubtedly an indicator of welfare. For instance, wealth provides insurance against periods of income loss. Put differently, a reduction in the flow of income can be compensated for through some depletion of the stock of wealth. The value of one's physical property tends to reflect physical space in which to live, which in turn often correlates with well-being. These relationships apply at the level of the individual, the household and a country. Important relationships exist between wealth and consumption. More wealth is associated with a lower need to invest current income and hence with an improved capacity to consume. Traditional welfare measures based on income alone fail to take into account these key distinctions. Wealth can be seen to include, apart from physical and financial capital, human capital and a sustained ecosystem. Poor human capital and a degraded ecosystem limit the extent to which income improvements can lead to a better quality of life. The SNA does in fact incorporate substantial wealth considerations. However, implementation of this in the actual accounting practices of countries is limited. The Quarterly Bulletin of the South African Reserve Bank reports subtractions from and additions to the physical assets held by firms and the government through the categories gross fixed capital consumption, meaning the depreciation of physical assets, and gross fixed capital formation, or investments in these assets. However, the physical assets of households, in particular dwellings, do not receive this attention. The total level of financial assets in the country is reflected in the Quarterly Bulletin, but not the total level of physical assets, let alone the total levels of human capital or ecological assets. SSF (2009: 91) advocate using consumption variables, such as net domestic product (NDP) and net national disposable income, as opposed to pure income variables, so that the need to invest in capital formation (possibly due to limited inheritances of capital assets) is properly taken into account. The values for these more consumption-oriented variables can in fact be imputed from existing national accounts figures (as they appear in South Africa's Quarterly Bulletin, for instance), though they are not explicitly present.

(d) Non-market production. As discussed earlier, goods such as vegetables grown by a household for its own consumption and services such as child-care that is not paid for, are not accounted for within actual national accounts. Empirical analysis has shown that if non-market production were included in national accounts, figures would rise substantially. For instance, the GDP of the USA would increase by around 30% (SSF, 2009: 130). This has two far-reaching implications for comparisons across time and space. Firstly, the exclusion of non-market production results in an exaggeration of the income (and implicitly, welfare) gaps between developed and developing countries given that the latter tend to have more production outside the monetised market. Secondly, income improvements seen over time would be an exaggeration insofar as these improvements would partly be a reflection of increased marketisation of production, not an increase in production.

(e) Inefficiencies in government services. GDP always includes the value of government services. These services are valued at the cost to the government. Government failure, in the sense of inefficiencies in the provision of public services, means that GDP per capita will overstate the level of welfare of citizens. Put differently, the same amount of money represents less human welfare if associated with public services, compared to if it were associated with goods and services in the private market. This point can be illustrated through reference to the voucher system in education, an education financing system widely debated though infrequently implemented where (according to the most market-oriented definition) government issues vouchers to households which are redeemable for education services in a private education market where government's role is reduced to that of quality controller. This radical proposal is often associated with Milton Friedman (1955), Nobel laureate for his work in consumption analysis (amongst other areas). Voucher system proponents point to its ability to reduce government inefficiency through the reduction of monopoly effects. Essentially, if a school does not serve a community well then an entrepreneur can establish a better school and parents can switch their human capital investments to the new school. Such efficiency-enhancing dynamics are not possible within a traditional and essentially monopolistic public schooling system such as the one existing in South Africa. Economic theory would suggest that a school voucher worth a specific amount would represent a higher level of welfare than an equivalent level of spending on a pupil by the government through a traditional public schooling system. Estimating the bias to GDP per capita comparisons across countries and over time associated with different levels of government efficiency would be difficult and seems not to have occurred. The key problem would be attaching monetary values to government inefficiencies, for instance the loss incurred by households as a result of not enjoying the freedoms of a voucher system.

(f) Institutional failure and defensive spending. Defensive spending is undertaken by households or firms or the government to deal with institutional weaknesses, relative to some standard of what is ideal. For instance, poor performance of institutions associated with law and order, combined with a poor level of social cohesion and large social inequalities, typically leads to large defensive expenditures by all three economic roleplayers in the form of replacement of stolen goods, large investments in the securing of properties, and prisons. Poor public transport systems are also typically considered a cause for high defensive spending on, for instance, private vehicles beyond what can be considered ideal. South Africa's level of defensive spending can be considered especially high with respect to both security and transport. Welfare in another country with the same GDP per capita but with greater security and better public transport, would be better, even in terms of simple consumption criteria. Within the South African education sector defensive spending has been put forward as a partial justification for teacher salaries that are higher than those in other middle income countries. South African teachers must spend more on private transport than their peers in other countries, for example. SSF (2009: 103) suggest ways of accounting for defensive spending. However, these methods are not easy, nor do they cover a wide range of defensive spending types.

(g) Leisure time. More leisure, or more time not spent producing out of choice, is considered to be positively correlated with welfare. In economics, the value of a person's leisure time is generally taken to be the opportunity cost of not working, in other words the monetary rate for leisure equals one's wage rate. Proper analysis of the levels of leisure enjoyed in a society require data on earnings and on time use. To deal with the latter, developed countries have increasingly made use of time use surveys. One such survey has occurred in South Africa, in 2000 (Statistics South Africa, 2001). Cross-country comparisons are typically made complex, though not entirely impossible, due to differences in the design of time use surveys and questions around the definition of leisure. Just one example of a number of complexities is how to deal with what can be referred to as imposed leisure. How should leisure that is not chosen, for instance due to labour laws or social norms, be valued? Should a country that makes it especially difficult for women to work be considered a country with an enhanced level of welfare, all else being equal? Presumably not. But this would be a relatively extreme case. SSF (2009: 135) explain how in general the use of typical income measures, without any consideration of leisure, to compare welfare across countries has tended to over-state the welfare of countries, such as the USA, where considerably less leisure is enjoyed than in other countries, such as France.

(h) The cost of risk. SSF (2009: 123) discuss the trend in countries such as France for saving to become more private and less public over time. This implies greater risk for individuals who, for instance, must increasingly take decisions about where to place their retirement savings as greater use is made of a diversity of investment options. Typically, or ideally, this trend is associated with spending on insurance against the risk that particular investment options will perform poorly or even collapse. In some ways, this is defensive spending and analogous to spending more on private security due to a deterioration in publicly run security services. What SSF do not discuss is the possibility of an opposing effect whereby the shift towards more private saving improves, for at least a part of society, investment choices, retirement earnings and hence welfare.

(i) Income inequality. A basic principle in welfare economics is that income inequality undermines society's overall welfare. In other words, if two societies enjoy an equal average level of GDP per capita and are similar in all respects except for the fact that income is distributed more unequally in the first society than the second society, then the overall welfare of the second society can be considered superior. Within-country inequality with respect to basic income measures are commonly referred to, for instance by means of the Gini coefficient in the UNDP's Human Development Report (UNDP, 2010: 152). SSF (2009: 135) point to the need to consider inequalities with respect to better measures of welfare, in particular measures that take into account real consumption and the amount of leisure enjoyed. One could go further and consider the presence of institutional factors that strengthen income inequality. The degree of democracy in a country can be expected to be a major contributor to income equality. Recent cross-country data analysis in fact supports this hypothesis. The political scientists Reuveny and Li (2003) find that degree of democracy promotes income equality when controlling for economic openness. They argue that it is essential to use economic openness as a conditioning variable. The economists Mueller and Stratman (2003) find that the level of voter participation in democracies promotes income equality. Both studies use a panel of country-level values spanning several decades which includes both developing and developed countries. Mueller and Stratman (2003) moreover use an instrumental variable approach to test the direction of causality. But there are costs associated with democracy. Reuveny and Li (2003) find that the size of the public sector is larger in more democratic societies, which suggests that these societies operate more inefficiently and thus experience a lower average level of income. On a theoretical level, Arrow's widely quoted impossibility theorem implies, in a nutshell, that elections where voters must choose between three or more options result in Pareto inefficient results. Sen (1999: 251) argues that Arrow's theorem is frequently misinterpreted to produce overly pessimistic conclusions around the efficiency of democracy. He argues that if a democratic system is characterised by sufficient information, in the form of, for instance, informed public debates, the contradictions implied by Arrow's theorem do not pose an insurmountable problem.

Topics (a) to (i) above all carry the implicit or explicit assumption that one should strive towards having more nuanced monetary measures of welfare, where the nuancing is to a large extent a matter of measuring what people consume, including leisure 'consumed'. Moreover, the assumption is that more consumption equals greater welfare. The next two topics move beyond this framework, yet remain narrow in the sense that they exclude the ecology and ecological sustainability (how ecology is brought into the equation is the subject of section 3.5 below). The two topics discussed here can be labelled (j) subjective wellness and (k) capabilities enjoyed due to one's objective circumstances.

(j) Subjective wellness. Many items that are important for human welfare are not traded and hence carry no explicit monetary value. An example would be additional life years. Before discussing non-monetary valuations of welfare, it is pertinent to consider how non-traded items such as life years have been assigned monetary value in economics. SSF (2009: 154) see as seminal in this regard the work of Gary Becker, often considered the founder of human capital theory and 1992 Nobel laureate for his application of microeconomic theory to non-market behaviour. Using welfare economics tools such as consumer theory and indifference sets, tools designed to assign monetary values to non-traded items, Becker and others have attached monetary value to years of human life and found that if a person's total consumption over his life (including 'consumption' of life itself) is one's key measure of welfare, then standard measures of income improvements in GDP per capita for the period 1960 to 2000 under-state improvements by about a third. In other words, standard measures have failed to take into account substantial human welfare improvements associated with greater life expectancy (Becker, Philipson and Soares, 2005: 279).

However, SSF (2009: 154) conclude that tools for attaching monetary values to non-traded items that are significant for welfare are too fraught with inaccuracies and unnecessary complications to be useful for policymakers. Rather, the use of non-monetary measures for these items has proven to offer better opportunities for valuing welfare. Specifically, subjective valuations of happiness have increasingly been used in economics. Whilst these subjective variables are not based on people's willingness to pay, a key criterion in consumer theory, they have been found to perform well in several economic models. Often three basic measures of subjective well-being are distinguished from each other. Firstly, life satisfaction is a person's valuation of whether his aims in life have been fulfilled up to that point in life. Secondly, positive affect is the presence of positive emotions such as enjoyment. Thirdly, negative affect is the absence of negative emotions such as anxiety or the experience of physical pain. These three subjective variables have been found to be less correlated with each other than one may intuitively think, which has led a number of analysts to emphasise considering all three variables simultaneously in any analysis of well-being (SSF, 2009: 146).

Two international surveys that probe subjective wellness stand out, the World Values Survey and the Gallup World Poll. The World Values Survey involves face-to-face interviews with a representative sample of all adults in a country. It is run every five years and sample sizes vary between 1,000 and 1,500 individuals. Up to 73 countries participate per survey wave. South Africa has participated in the last four waves of the survey, with the South African sample rising to around 3,000 respondents in the last wave of 2007[33]. The Gallup World Poll has occurred annually in recent years and also involves face-to-face interviews. It covers over 130 countries. South Africa has participated since 2006, with a different sample of around 1,000 individuals every year[34]. Gallup sells its data, whilst the World Values Survey data are available as free downloads.

Average subjective wellness values per country have increasingly been used for international comparisons. These comparisons have been the subject of lively conceptual and methodological debate. Much of the debate has centred around the so-called Easterlin paradox, or the finding that although subjective wellness and income are relatively well correlated within countries, between countries they are not. Essentially people in rich countries are not systematically happier than people in poor countries even though the rich in one country tend to be happier than the poor in the same country. This observation has led to the conclusion that people's happiness is largely based on their material welfare relative to others directly observed in the same society. An important and recent critique of the Easterlin paradox is provided by Sacks, Stevenson and Wolfers (2010), who argue that the apparently poor correlation between subjective wellness and income at the cross-country level was largely a symptom of inappropriate functional form and errors of measurement. After correcting for this, these analysts obtain a correlation coefficient of 0.82 between logged income per capita and subjective wellness across 131 countries, using Gallup life satisfaction values for the latter variable[35]. South Africa falls below the regression line, meaning South Africans are less satisfied than what income would predict[36]. This is to be expected given the large gap between mean and median income in South Africa caused by the country's exceptionally high income inequality.

Debate around whether the Easterlin paradox exists has been lively since Richard Easterlin first proposed it in 1974. However, insofar as a recent stocktaking of the literature by Layard, Clark and Senik (2012: 66) can be trusted to offer a balanced verdict, it appears that evidence does not support the existence of the Easterlin paradox. Economic growth, it seems, is good for happiness.

Analyses where subjective wellness variables replace the income variables in growth models such as those discussed in section 4 seem not to exist. However, a close proxy for this is offered by Di Tella, MacCulloch and Oswald (2003: 812), who in an article titled 'The macroeconomics of happiness' regress life satisfaction on a range of explanatory variables, including education variables and income, at the level of individuals across several European countries, with dummy variables for countries. They find that having continued with education beyond age 19 is positively associated with greater life satisfaction after controlling for income. The magnitude of this education effect is around a quarter of the effect of not being unemployed and a half of being divorced – unemployment and divorce are found to be amongst the largest determinants of unhappiness. This analysis suggests that there are reasons beyond the often cited income reasons for expanding post-school education.

---

[33] See http://www.worldvaluessurvey.org. The description of the WVS and the Gallup World Poll provided here reflects the situation in 2011.
[34] See http://www.gallup.com.
[35] Sacks, Stevenson and Wolfers, 2010: 43 (see Figure 4).
[36] Sacks, Stevenson and Wolfers, 2010: 42 (see Figure 3).

Subjective measures of wellness have largely been a topic of debate amongst academics and not policymakers. One notable exception is Bhutan, where the king introduced the term 'gross happiness index', or GHI, in 1972. This indicator has since then become an important element in the national planning process of Bhutan and in some respects Bhutan has come to be seen amongst analysts from a wide range of countries as a laboratory for the use of this kind of indicator. Bhutan hosts regular international conferences on the topic[37]. Bhutan's official GHI is in fact not a standardised time series of single values, but rather an umbrella term under which a number of variables are studied for planning purposes in a relatively flexible manner.

(k) Capabilities enjoyed due to one's objective circumstances. The subjective measures discussed above are naturally limited by their very subjectivity. One criticism that has been made is that people tend to be pessimistic and under-value improvements in their circumstances over time (SSF, 2009: 152). One way of escaping the limitations of both income and subjective variables is to focus on the objective circumstances of people which according to experts or broad consensus are important for well-being. A key expert associated with the emergence of this approach is Amartya Sen, 1998 Nobel laureate for his contributions to welfare economics. Sen's 1999 book *Development as freedom* encapsulates many his proposals for valuing welfare. Sen's work focuses largely on what to aim for in development and less on the precise mechanisms for attaining the aims though on certain points, such as the vital role of a free market, Sen is emphatic. Sen's (1999: 38) basic thesis is that five classes of 'instrumental freedoms' enjoyed by people should all be strongly promoted in the development process, both because they are a means to greater welfare but also because these freedoms are constituents of welfare and hence worthwhile ends in themselves. The five types of freedoms are discussed below.

- **Political freedoms.** Democratic freedoms are important, Sen argues. India and China are contrasted to make the point. A lack of political freedoms in the case of China makes this country less able to make the necessary economic policy adaptations in times of crisis. The fact that democratic India has not suffered a major famine since independence in 1947 is contrasted with China's major 1958 to 1961 famine. The other four types of freedom discussed below can be considered the outcomes of a successful democracy even if to some extent they are realisable in the absence of democracy.

- **Economic freedoms.** These freedoms allow people to transact economically in a fair and predictable manner through a well functioning market, where the state promotes the enforcement of contracts and law and order generally.

- **Social opportunities.** Under this heading Sen is referring above all to publicly organised education and health services that promote the basic capabilities people need, for instance, to participate in democratic processes and the market.

- **Transparency guarantees.** This is perhaps the least predictable of Sen's five categories of freedoms and it is interesting that it should be prioritised as it is. Transparency guarantees provide people with access to the information that they need and counteract incorrect information. The implied policy priorities are government mechanisms to protect and promote freedom of expression and to

[37] See http://www.bhutanstudies.org.bt.

allow access to needed government information. However, the reach of this category is wider than government policies and encompasses sufficient levels of honesty and trust within society. This echoes the notion of social capital of sociologists such as James Coleman (most known in economics of education circles for his groundbreaking 1966 Coleman Report, arguably the first ever fully-fledged education production function). Transparency guarantees act as a bulwark against ‘corruption, financial irresponsibility and under-hand dealings’ (Sen, 1999: 40).

- **Protective security.** Here Sen is referring to government initiated safety nets to assist people during temporary crises relating, for instance, to unemployment and spikes in food prices.

Sen’s development framework has been widely embraced by policymakers[38]. The intended comprehensiveness of Sen’s framework inevitably invites criticisms around what has been omitted. An omission that seems critical from a South African viewpoint is the absence of any explicit treatment of trade unions in *Development as freedom,* despite their important role in many countries. For instance, do trade unions cause economic ‘unfreedoms’ insofar as they promote minimum wages above the market clearing level? If this is true (as one might suspect following Banerjee, Galiani, Levinsohn, McLaren and Woolard (2007) or Kingdon and Knight (2007)), would limiting the rights of workers to act collectively not constitute another economic unfreedom? In other words, are there not difficult contradictions within Sen’s framework that should be acknowledged? Whilst Sen may be silent on trade unions, on unemployment he is very clear. In comparing the United States and Europe, he argues that an excessive focus on income as a measure of welfare has obscured the detrimental effects of long-term unemployment in Europe, which has historically been much higher than in the United States. Whilst extensive income transfer mechanisms in Europe deal with the problem on the income side, the failure of Europe to deal more decisively with the underlying unemployment problem represents a serious crisis of under-development, according to Sen (1999: 94). Similar arguments could be made in the case of South Africa.

On one level, Sen disagrees with the need to inform policymakers using the kinds of growth models dealt with in this dissertation. The following from Sen (1999: 5) is revealing:

> …the opportunity to receive basic education or health care… are among the *constituent components* of development. Their relevance for development does not have to be freshly established through their indirect contribution to the growth of GNP or to the promotion of industrialization. As it happens, these freedoms and rights are *also* very effective in contributing to economic progress…

That basic education should be an end in itself and carries benefits that are not related to income is something Adam Smith supports when he argues in his *Lectures on jurisprudence* with respect to the uneducated child: ‘When he is grown up he has no ideas with which he can amuse himself.’ (Smith, 2010: 460) However, Sen is not

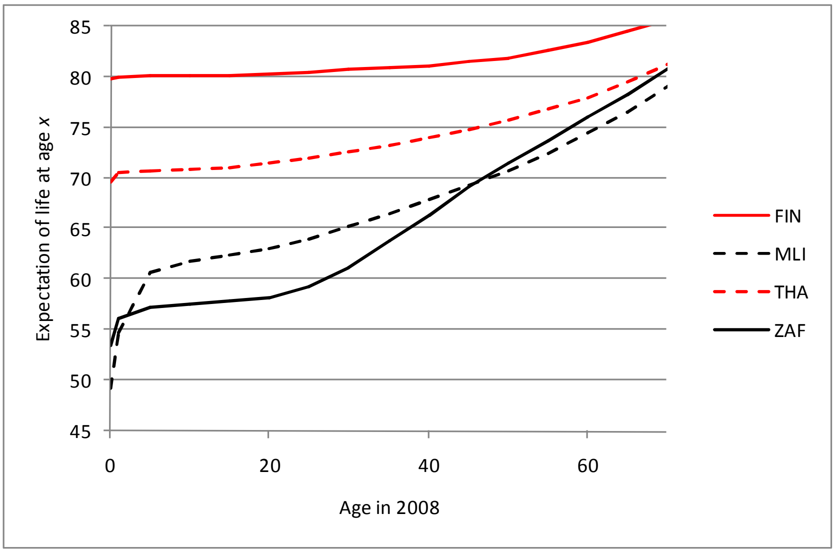
---

38 ‘The Guardian Profile: Amartya Sen’ at http://www.theguardian.com/books/2001/mar/31/society.politics. The South African Minister of Finance, Trevor Manuel, made reference to Sen in his 2003 budget speech (http://www.treasury.gov.za/documents/national%20budget/2003/speech/speech.pdf).

impervious to the importance of the causal link between education and income. He in fact emphasises that improved income per capita in developing countries is an important prerequisite for the realisation of many freedoms (Sen, 1999: 25). Moreover, Sen emphasises that human capital development should contribute towards human capabilities. The key omission in this regard in Sen (1999) is perhaps how easy it is for educational activities *not* to translate into improved human capabilities, in other words how easily paying attention to only educational activities, and not educational outcomes, can provide an illusion of progress. This omission in Sen is likely to be linked to the fact that much of the work that has emphasised poor educational outcomes, for instance Van der Berg (2007) in the case of South Africa, post-dates much of Sen's work.

Sen has played an important role in the establishment of the UNDP's Human Development Report series, which started in 1990 and has produced a report virtually every year since then. A well-known element of these reports is the Human Development Index (HDI), an indicator that combines income per capita, the level of pre-tertiary enrolments, adult literacy levels and life expectancy. Its aim is largely to act as a counterweight to simple income indicators. In particular, the indicator is intended to highlight development problems in countries whose performance with respect to income is better than their performance with respect to the other HDI components. South Africa falls into this group of countries, mainly due to its exceptionally low life expectancy, itself an outcome of the HIV/AIDS pandemic. Despite widespread acceptance and media coverage, the HDI has been severely criticised. It deals only with the second and third categories of freedoms discussed above, and even with respect to these categories the coverage is partial. The HDI is thus far from being an all-encompassing development indicator. Moreover, the weights attached to its various components are largely subjective and not empirically informed. Sen was in fact initially opposed to the notion of collapsing several development variables into a composite indicator. With respect to the HDI formula, SSF (2009: 56) point out that 'adding the logarithm of per capita GDP to the level of life expectancy... implicitly values an additional year of life expectancy in the United States as worth 20 times an additional year of life in India'. The precipitous drop in life expectancy at birth in countries such as South Africa and Botswana since around 1995 and the consequent decline in the national HDI values provide a further indication of the problem with the HDI methodology. The patterns of mortality brought about by HIV/AIDS make the interpretation of the life expectancy at birth indicator even more difficult than was previously the case. The problem is illustrated in Figure 15 below. Life expectancy at birth in 2008 was 49 in Mali and 54 in South Africa. Yet in many ways children in Mali were better off in the sense that from about age five, life expectancy was well above that in South Africa. Mortality in Mali was concentrated during the first couple of years of life. Above age 45, however, the situation for South Africans surpasses that of Malians. Considering life expectancy only at birth in isolation from patterns seen at other ages can be deceptive. (The curves for Finland and Thailand have been included in the graph to illustrate more or less where the best middle and high income country levels lie.) Clearly there is a risk that over-interpretation of the HDI will obscure important background distinctions. In fact, SSF repeatedly stress the risks inherent in trying to collapse too much information within one indicator and prefer a more nuanced approach of considering several key indicators simultaneously.

Figure 15: Age-specific life expectancy



Source: World Health Organization (2011).
Note: The ISO country codes refer to Finland, Mali, Thailand and South Africa.

Some indication has been provided above that changing the dependent variable in a growth regression from income or income growth to what some may consider a more relevant measure of human welfare, does not change the fundamental finding that education is a key variable for human development and the attainment of greater levels of welfare. Importantly, this finding holds even when a dependent variable that is conceptually very different from income or consumption, namely subjective life satisfaction, is used. The relevance of the foregoing discussion for education planners relates not so much to whether education contributes towards development, but to how to educate the young. Undoubtedly, a strong focus on income, as opposed to wellness or people's objective circumstances, is a feature not just of economic models, but also society and governments in general. The discussion suggests strongly that some of this focus on income may be misplaced and should instead shift towards, for instance, questions around Sen's enabling freedoms. This has implications for the design of school curricula and textbooks, for instance. To a fairly large degree, these implications are considered by curriculum designers, though many have argued that this has been done too superficially and without sufficient recognition of realities within schools. For instance, Harley and Wedekind (2004) argue that in South Africa, whilst pro-development values have been promoted in the post-apartheid school curriculum, insufficient attention to pedagogical method has resulted in a situation where pupils do not develop fundamental skills, without which their future capabilities are limited. The next sub-section, which deals with sustainable development, will highlight further and very formidable challenges for those shaping what gets taught in schools.

3.5 Growth versus sustainable development

The term 'sustainable development' is usually traced back to the UN's Brundtland Report of 1987, which took stock of environmental degradation across the world and discussed the risks that this posed for future welfare. Sustainable development, the report explained, is development which will 'ensure that it meets the needs of the present without compromising the ability of future generations to meet their own needs' (United Nations, 1987: 24).

The absence of ecological considerations in the models of development presented in sections 2 and 4 would seem to be a significant gap. Tim Jackson, an 'economics commissioner' in the UK's Blair government, and whose arguments are discussed below, has said 'economists... are schooled in a language that rarely even refers to natural resources or ecological limits... Economics – and macro-economics in particular – is ecologically illiterate' (Jackson, 2010: 123). Ecological factors have until fairly recently seemed sufficiently minor to allow economics to produce valuable policy advice without taking these factors into account. However, the realities of the greenhouse effect and the related matter of climate change have made it increasingly important to bring ecological factors to bear on economic models, especially models dealing with growth.

When did climate change emerge as a major human development issue? Though scientists had speculated about the existence and nature of anthropogenic climate change since the late 1800s, the point at which sufficient consensus was reached amongst scientists is arguably a 1985 conference held in Villach, Austria, which triggered a series of events that led to the establishment of the Intergovernmental Panel on Climate Change (IPCC) in 1988 (Agrawala, 1998). Thus 1985 can be considered the point at which economists were presented with the challenge of explaining, in economic terms, an ecological shift that was far greater and more imminent than anything that had emerged previously. Importantly, widespread public concern about climate change, and hence climate change as an election issue for voters, at least in some countries, emerged much later. Media analyses have identified a critical surge in public concern in the 2004 to 2006 period, a period during which media coverage of the issue quadrupled and after which coverage appeared to remain high. These timeframes form an important backdrop to the fairly rapid developments that have occurred in economics in response to the climate change question.

The response within the field of economics has its roots in methodologies developed for the sub-field of environmental economics, beginning in the 1950s. An intellectual history of this sub-field, by Pearce (2002), describes an important tension between environmental economists, economists who attempted to incorporate environmental factors such as pollution within an economic framework, and ecological economists, a 'broad church' of largely natural scientists who, dissatisfied with the slow progress made by the economists, sought new ways of combining ecological models and economic thinking. A number of factors contributed to a state of affairs where ecological issues within economics were confined to one sub-field, environmental economics, and, even within this sub-field, were dealt with in a rather limited manner. One factor was simply insufficient contact across disciplines, which delayed the realisation within economics of the magnitude of the ecological effects of modern industrial economies that had become clear to the natural scientists. Moreover, what is described as marginalism in economics, or a specialisation in how marginal or relatively small effects change a system, has made it difficult for economists to take into account the large ecological and hence also economic effects implied by the climate change literature (DeCanio, 2003: 3). Put differently, economists had not seen the need to consider large discontinuities in the ecology within which the economy is situated. Nicholas Stern (2006b: 25), an economist whose work is discussed below, describes the climate change discontinuity problem as 'market failure on the grandest scale the world has ever seen'.

The way environmental economics has traditionally dealt with the problem of internalising an externality cost brought about by environmental degradation is to specify that the firm responsible for the pollution should be taxed an amount corresponding to the marginal cost of the social damage caused by the pollution (Hyman, 1999: 102). This will result in both a decline in the demand for the product associated with the pollution, and hence less pollution, as well as additional tax revenue that can be used to assist those who fall victim to the pollution. This framework is limiting in terms of both time and space if one attempts to address climate change within it. Firstly, given that the effects of the greenhouse gases causing climate change are felt at least a couple of generations after emissions have occurred, important inter-generational equity questions must be addressed (Stern, 2006a: 4). Secondly, given the global nature of the climate change problem, any policy solutions must involve all the world's government and are thus complex and include important gaming and free rider dynamics between countries. The original theory of environmental economics, which assumed damages occurring within one generation and one country, is clearly an insufficient basis for climate change policy analysis.

Initial moves towards accommodating climate change within economics stimulated intense debate around, above all, two specific theoretical issues. One was how to model welfare trade-offs between several consecutive generations in a way that might inform what policy positions to take presently. The other was how to model trade-offs between different kinds of capital assets, including the ecology as an asset, over an extended period of time. In both of these debates, the question of whether to relax the strict welfare economics requirement that all welfare be measured in monetary terms has been a central concern. On the matter of inter-generational trade-offs, Pearce (2002: 77), in referring to future generations, considers that 'it is not clear whether nonexistent beings can be said to have any rights at all'. He is probably being provocative here. Yet intuitive notions of future people are clearly not sufficient for policy analysis in a context where policy decisions, or their absence, now can have potentially devastating impacts on future generations in ways that would have been unimaginable just a few decades ago. For example, if inequities between societies at the present time are more or less accepted by policymakers and voters, why should inequities over time not be more or less accepted? Technically, various discount rates, including ones based on the world real interest rate, have been used in order to express the welfare of future people in current terms. The real interest rate has generally been discarded as a feasible discount rate as it results in a valuation of future people that is far lower than what current ethical systems would regard as proper. At the same time, zero discount rates result in scenarios where any economic activity today that harms people, say, three thousand years in the future becomes treated as if it were causing the same harm today (DeCanio, 2003: 58). On the matter of trade-offs between assets expressed in monetary terms, Pearce (2002: 65) argues 'Probably the largest, and most controversial, research effort in environmental economics has in fact been devoted to [the] issue of valuing nonmarketed asset change'. The valuing of assets is central to the economic approach to considering the welfare of future generations, because future welfare is to a large extent based on the total capital assets that are passed on to future people. Stern (2006b: 42) expresses this notion as follows and includes a reference to human capital assets:

> …the principle [of sustainable development] need not imply that the whole natural environment and endowment of resources should be preserved by this generation for the next generation in a form exactly as received from the previous generation. The capital stock passed on to the next generation consists of many things, mostly in the form of stocks covering, for example, education, health, capital equipment, buildings, natural resources, the natural environment, etc. The standard of living available to the next generation depends on the whole collection of stocks. A decline in one of them, say copper, might be compensated by another stock, say education or infrastructure, which has increased.

As attractive as this notion may seem, it has limitations. For instance, in a context of uncertainty, it is not possible to establish an 'exchange rate' between, say, education and the burden of greenhouse gases in the atmosphere.

Two widely known and interesting responses to the challenge of adapting growth policies to deal with climate change, one by an economist and another by an ecological economist, are discussed below, partly with a view to arriving at some implications for the education sector. The sources of the two responses, Nicholas Stern and Tim Jackson, both produced important reports for the Blair government in Britain. The British government has been especially prominent, at least within the English-speaking world, as an investor in policy research relating to climate change. In fact, it has been argued that Margaret Thatcher, whilst British Prime Minister, was prescient in her attempts to push climate change up the global policy agenda (Thatcher was trained as a natural scientist). She was unsuccessful in this attempt (Flannery, 2005: 247).

In Stern's (2006b) *The economics of climate change*, also known as the Stern Review, the question of the degree to which future welfare should be discounted within current policy considerations is largely avoided. Instead, limiting average temperature increases to around three degrees Celsius is considered something worth devoting considerable policy effort towards, whilst at the same time greater restrictions on the extent of climate change are considered to be unattainable. These *a priori* criteria are based on assessments by natural scientists of the likely weather irregularities, rainfall disruptions and sea level rises associated with different degrees of warming. Integrated assessment models, or IAMs, are used for the analysis. These are models that emerged in the 1980s and have focussed on combining macroeconomic and ecological forecasting. Stern (2006b) makes use of an IAM that takes into consideration the uncertainty and probability ranges inherent in climate prediction. For this, Monte Carlo simulations are required. This allows the impact of current policy decisions to be seen in terms of risk reduction. Stern lets the modelling deal with two questions fairly separately. The first question is the likely impact of climate change on GDP per capita if no special policy measures are taken, in other words in a business as usual (BAU) scenario. Climate change impacts on income occur mainly through disruptions in production, particularly within agriculture, and adaptation, which could involve migration and the construction of infrastructure such as canals and levees. Adaptation could impact negatively or positively on income (the contradictory nature of defensive expenditure was discussed in section 3.4). One plausible model (Stern, 2006b: 157) finds that there is a 90% probability that a BAU scenario would result in a drop in GDP per capita of between 4% and 35% by the year 2200, relative to a situation in which there was no climate change. Initial losses are relatively low: a 1% loss in global GDP per capita by 2050 and about a 5% loss by 2100. In the BAU scenario, the average global temperature by 2200 is four degrees Celsius above pre-industrial levels. Although Stern (2006b) does not always make this point very clear, the base scenario with no climate change has sufficient income growth in it to ensure that all the scenarios with climate change retain positive economic growth, at least at the global level.

The second question examined in Stern (2006b) is the impact on GDP per capita of likely mitigation policies implemented in the near future to limit temperature increases to three degrees above pre-industrial levels. Three types of policy measures are considered in the model, relating to carbon pricing, technology innovation and behavioural change. Here the cost is found to be a worsening drop in annual GDP growth, relative to the BAU scenario with no policy measures, reaching a reduction in annual growth of 1 percentage point by 2050. However, the costs of the policy measures would be small relative to the economic growth expected to occur during the first half of the twenty-first century: overall growth of around 200% in OECD countries and 400% in non-OECD countries. Assuming a linear phasing in of the GDP costs associated with the mitigation policies, the total loss in GDP by 2050 would be just 20%. Even with population increases, average GDP per capita in 2050 would be at least twice what it was in 2000. (Surprisingly, Stern leaves out these final calculations relating to total mitigation costs and population trends.) The finding that substantially increased income per capita can be combined with reduced ecological impacts is something that has attracted criticism, as will be seen below. What is not made clear by Stern is why the costs of climate change and the costs of mitigation policies are examined within separate models. The IAM used by Stern, the PAGE2002 model, is said to be able to model the two simultaneously[39]. The explanation could lie in the different time horizons of the two analyses, which is related to the substantial lags between greenhouse gas emissions and the economic impact of those emissions. The lag is between 50 and 100 years, meaning for instance that climate change experienced in 2010 would largely be an outcome of the concentration of greenhouse gases in the atmosphere that had been reached in around 1940. Put differently, if anthropogenic greenhouse gas emissions had suddenly stopped in 1940, climate change in 2010 would not differ much from what was actually experienced in 2010. One problem with Stern's separation of the analysis into two models is that it is not possible to see to what extent the GDP per capita situation in 2200 (or before then) is improved, relative to the BAU scenario, if the envisaged mitigation policies are carried through.

Tim Jackson (2009a, 2009b) titles both his book and the government report he authored *Prosperity without growth*. Jackson's arguments, like those of his several intellectual predecessors, question the policy relevance of the models discussed in sections 2 and 4 rather fundamentally. Jackson's basic disagreement with an approach such as Stern's rests on the conclusion that the ratio of greenhouse gas emissions to each dollar of GDP cannot be reduced to the degree envisaged by Stern. Put differently, an increase of, say, 300% in income is incompatible with a reduction in greenhouse gas emissions, something that Stern implicitly states is possible (though Stern's logic in this regard is not clearly stated). The argument that GDP per capita

[39] See http://unfccc.int/adaptation/nairobi_work_programme/knowledge_resources_and_publications/items/5447.php. PAGE stands for Policy Analysis for the Greenhouse Effect.

can increase whilst emissions per capita decrease is referred to as the 'decoupling' argument or, more pejoratively, as the 'angelising' of GDP. Essentially GDP becomes to a large degree decoupled from physical production as we currently understand it and becomes more non-physical, depending increasingly on services and information. Though Jackson does not provide a comprehensive case against decoupling, he does make the important point that what may appear as decoupling in the last few decades in the form of a decreasing ratio of tons of greenhouse gas emissions to each dollar of GDP in rich countries is in fact deceptive due to the exclusion of international trade from the calculations. Reports produced by countries within the emissions reporting systems of the UN focus on emissions associated with production and not consumption occurring within national borders. As a result, these reports have reflected declining emissions for countries such as the UK, though if one takes into account the emissions associated with consumption in the UK, total emissions and emissions per capita have risen (Jackson, 2009a: 73). Essentially, the British have increasingly imported those products associated with high levels of emissions.

Jackson argues that policy measures beyond those advocated by Stern need to be implemented if the desired reductions in emissions are to be brought about. Specifically, he places a stronger emphasis on changing the behaviour of consumers and emphasises that existing mixed signals of rich country governments (increase your production and consumption but be concerned about the environment) are problematic. In addition, Jackson sees restrictions to the length of the working week as necessary in order to control consumption. In this latter regard, he relies on the work of Victor and Rosenbluth (2007), whose analysis of Canadian data suggests that a reduction in the working week of just 3%, combined with a number of other policy measures, can bring about the desired reduction in emissions associated with lower consumption. In effect, consumption is suppressed because income is suppressed. Jackson (2009a: 34) points to this work as the beginnings of an 'ecological macro-economics'. Part of the importance of this work, according to Jackson, is that it indicates that low or zero growth is compatible with political and economic stability, a matter that for Jackson represents an area of risk within his larger argument. With respect to investments in cleaner sources of energy, Jackson also highlights risks that were not considered in Stern (2006b). Here Jackson borrows from D'Alessandro, Luzzati and Morroni (2008), who use a theoretical model to demonstrate the existence of a 'sustainability window' during which sufficient investments in cleaner energy must be made. In a nutshell, growth spurred by unrestricted fossil fuel use has created a stock of financial capital which must be invested in cleaner energy. However, as the switch to cleaner energy occurs, growth will decline and savings will become depleted. There is a risk that insufficient savings will remain for the required levels of investment. Jackson (2009) thus sees very active stimulation of the savings rate as an important part of the policy response to climate change.

Jackson's policy recommendations are largely aimed at rich countries. However, they have relevance for policymakers in developing countries too for a number of reasons. Firstly, they imply that development involves a gradual transition from a focus on income growth towards a focus on other development indicators and possibly a state of zero growth. This changes the textbook approach to development economics rather fundamentally. Secondly, a shift away from promoting consumption growth in rich countries would impact on export-oriented industries within developing countries.

Stiglitz, Sen and Fitoussi (2009: 70) lean towards the Jackson argument when they advocate switching some of the development attention away from GDP growth and towards indicators that reflect the economy's ecological sustainability. Of the many existing indicators available for the latter, two are preferred. They are the ecological footprint of the Global Footprint Network[40] and the WWF (2010) as well as the World Bank's adjusted net savings (ANS) indicator, for which annual values for 170 countries are available for the period since 1970 (Bolt, Matete and Clemens, 2002). This last indicator accounts for not only the capital stock of produced physical capital, but also the stock of human capital and the stock of natural subsoil assets (however it does not take into account atmospheric factors associated with climate change).

If, as Jackson argues, macroeconomics is ecologically illiterate, then both macroeconomics and ecological economics can be considered demographically illiterate. It is indeed surprising that in a context of a large increase in the world's population, from around three billion in 1950, to seven billion in 2010, to a projected figure of around nine billion in 2050, the matter of population policies receives very little attention in the climate change literature. For instance, there is little emphasis on ways in which population increase, given what we now know about climate change, may compromise growth. The silence with respect to population policies is partly attributable to the stigma associated with policy initiatives in recent times, from the eugenics movement in the early twentieth century to the coercive one child policy in China. However, other factors such as religion have also played an important role. The absence of analysis and policy debate in the area of demographics seems unfortunate. In the literature that has argued for a more proactive treatment of the subject, the documents arising out of the critical 1994 Cairo conference on population and development have been strongly criticised for their lack of focus on the numbers, policies and ecological impacts of population growth. The costs of this lack of policy attention may be high. Some analysis suggests that if unmet demand for contraception were eliminated, world population by 2100 would be two billion lower than expected. Some analysis has pointed to policies aimed at encouraging lower fertility on a voluntary basis as being cost-effective ways of tackling emissions (O'Neill, MacKellar and Lutz, 2005; Hodgson and Watkins, 1997).

Bangay and Blum (2010) provide a rare example of a discussion of the implications of climate change for education planning in developing countries. Clearly policy responses will be necessary to deal with the disruptions caused to the education system itself, in particular its physical infrastructure, and to the socio-economic conditions of pupils. These measures will be most necessary in developing countries, where the impacts of climate change are expected to be more damaging. Bangay and Blum (2010) use impacts of recent natural disasters on schooling systems in Asia, and the subsequent policy responses, to provide an indication of what responses should be planned for in future. What teachers teach in schools can improve both mitigation and adaptation efforts of societies. The general quality of schooling, as measured by indicators that received attention in section 3.3, will be an important determinant of any society's ability to adapt. Moreover, more education focussing on appropriate behavioural responses to climate change, such as more socially responsible consumption, will be needed, not just within a separate environmental education subject, but across all the subjects of the curriculum. Pupils also need to be equipped to engage in the complex ethical, economic and social debates surrounding climate

[40] See http://www.footprintnetwork.org.

change. Analysis of PISA data have pointed to a strong correlation between scientific knowledge and an understanding of environmental dynamics. This underscores the inter-connectedness of different education goals. A careful balance will need to be maintained to avoid a situation in which the promotion of appropriate behaviour becomes indoctrination, something that studies suggest has serious unintended consequences such as a culture of pretending to comply with prescribed norms whilst in fact compliance is minimal. What this means is that pupils need to be convinced certain behaviour is necessary because they understand the science behind the consequences of inappropriate behaviour.

The instilling of useful and desirable values in the schooling process is partly a question of what and how to teach across the whole curriculum and partly a question of how to design one specific subject, referred to as civic education (or civics) in the United States, citizenship education in the United Kingdom and life orientation in South Africa. The civic education curriculum tends to encapsulate the values that the school curriculum as a whole attempts to transmit. The United Kingdom and South Africa cases serve as fairly typical examples from democratic nations. In both countries, the civics curriculum emphasises social tolerance (or the avoidance of destructive values such as racial prejudice), understanding national and global political systems and involvement in community and political life. In South Africa (Department of Education, 2002), the emphasis on the latter skill, namely the skill of political involvement, is arguably under-developed relative to what is found in the UK curriculum. The UK curriculum emphasises an understanding of trade-offs and constraints in a way that would be familiar to economists. Specifically, 'Investigating ways in which rights can compete and conflict, and understanding that hard decisions have to be made to try to balance these' receives attention and pupils should study 'how economic decisions are made, including where public money comes from and who decides how it is spent' (United Kingdom, 2007: 29, 32).

Before some discussion of what this section suggests may be missing from civic education, it is instructive to consider whether civic education has been found to influence pupils in the intended manner. Galston (2001), with a focus on the United States, concludes that a long-reigning consensus that civic education has virtually no impact on behaviour has been successfully overturned by more recent empirical evidence pointing to specific impacts of this element in the school curriculum. Finkel and Ernst (2005), using data from South Africa in a multivariate regression analysis, find that the existence of civic education but also effective techniques in delivering this education do play a role in building political knowledge and instilling democratic values in pupils. Torney-Purta (2002), using a landmark International Association for the Evaluation of Educational Achievement (IEA) data collection of 1999 focussing on civic education in mainly OECD countries, also finds a positive impact. For the policymaker, these findings are encouraging insofar as they present clear possibilities for change through policy.

If there is a key ingredient missing in the typical civics curriculum, it is perhaps consumer education. How government failure, for instance in the form of inadequate public transport, leads to defensive spending and how the consumption inherent in this is inefficient and reduces welfare should form an important element of voter education, especially in countries where defensive spending is higher than it should be. One can extend the more typical concept of defensive spending explained in section 3.4 to encompass the 'conspicuous consumption' discussed by Jackson (2009: 99) whereby consumption in many ways becomes a signalling of status needed in a context of high social inequality. One consumes to avoid being mistaken for an uneducated member of society, or one whose capabilities have been curtailed. Understanding the sociology and psychology of conspicuous consumption seems to be an important prerequisite of the policy drive to reshape consumption in ways that promote ecological sustainability. Of course these aspects to civic education are relevant largely for the non-poor who are able to make a large range of consumption decisions, meaning most people in developed countries and the elites within developing countries. Yet they are important, especially if one considers the relationship between the consumption of the non-poor and the persistence of poverty, especially in a context of a changing climate.

3.6 Conclusion

The complexity of the debate around how to define and measure a country's progress is immense. The widely used inflation-adjusted GDP and GDP per capita measures enjoy respect, one suspects, partly because they have become institutionalised over several decades and their familiarity provides a certain assurance. But these measures are flawed, probably to a greater extent than most analysts who use and interpret them believe.

Before one even begins to consider matters of sustainability, two flaws that stand out with respect to the commonly used GDP measures of country development are that they ignore, firstly, income inequality and, secondly, defensive spending related largely to government failure. To illustrate the problem of income inequality, comparing South Africa, on the other hand, to Thailand and Ukraine, on the other, is instructive. Thailand's average GDP per capita was, in around 2010, more or less equal to that of South Africa in purchasing power terms, and that of Ukraine was around 15% lower than South Africa's. Yet if one considered everyone except for the top 20% of the population in terms of per capita income, Thailand and Ukraine perform substantially better than South Africa. The remaining 80% of the population is around 70% better off in both Thailand and Ukraine, compared to South Africa, in terms of average per capita income[41].

The distortions to income figures brought about by defensive spending are less easy to quantify. Government spending on actual defence where this has its roots in the government's inability to resolve internal divisions quantifies one part of the overall phenomenon of defensive spending. Israel, Algeria and Lebanon are all countries where government spending on defence exceeds 4% of GDP, largely because of unresolved internal conflicts. Unfortunately, the impact of defensive spending on welfare where this spending is more associated with inequality-related crime (South Africa), severe levels of bribery and corruption (Nigeria) and the need to comply with cumbersome and often illogical government regulations (Argentina) are not as easy to quantify.

[41] Online data-querying facility of World Bank (World Bank DataBank), consulted December 2012, used for income distribution. Data from the years 2008 to 2011 were used. Heston, Summers and Aten (2011) used for income figures.

Evidence on the ecological impacts of the last two hundred years or so of industrialisation has become increasingly available and believed and represents an immense challenge for the field of economics, which needs to develop new models explaining the relationships between economic activity, the ecology and sustainability for policymakers, including, to some extent, education policymakers. Nicholas Stern, the economist responsible for assessing responses to climate change in Britain's Blair government, has likened the consequences of not responding adequately to climate change to the world wars of the twentieth century (Stern, 2006: ii). Yet, as discussed previously, Stern has implicitly been accused of under-estimating these consequences. What seems far more prone to uncertainty than the predictions of the climatologists, given a business as usual scenario, are the predicted outcomes of policy decisions aimed at mitigating climate change or adapting to it. These uncertainties are an inevitable result of the complexity of the modelling that must occur. However, one suspects that a more fundamental uncertainty is not a measurement one, but an ethical one, around the degree to which a current generation should restrict its welfare in order to ensure that future generations can enjoy similar levels of welfare. In economic terms, the question is what discount rates should apply to future generations.

Existing GDP per capita measures, with all their flaws, will remain centrepieces in the measurement of country progress for a long time to come. However, it appears they will to an increasing degree be interpreted with caution and accompanied by a narrative that explains the caveats. Unfortunately, there does not appear to be much agreement on how to organise the accompanying narrative better. Where there does appear to be agreement is that sets of carefully interpreted indicators are better than over-emphasising any single indicator. But a good and widely accepted set of indicators with a good accompanying narrative, endorsed by bodies such as the World Bank and the UNDP, still seems several years from being realised.

Compared to the difficulties of attaching values to a country's level of overall development, measuring a country's level of educational development is relatively easy. Questions around whether to consider the quantity or quality of education are not so much a matter of fundamentally differing understandings as a matter of the available data. Human capital is undoubtedly about what people know and the skills they have. Around this there seems to be little disagreement. There also seems to be agreement now that years of education attained tends to predict the knowledge and skills of adults poorly, especially in cross-country comparisons. However, if one wants to cover a wide range of countries, the trade-off is currently one between having little historical data, if one uses average test scores from schooling systems, and having much historical data but of questionable relevance, if one uses attainment statistics.

4 A SELECTION OF EMPIRICAL GROWTH MODELS

4.1 Introduction

Three empirical cross-country growth models are discussed in depth below: (1) Mankiw, Romer and Weil's (1992) augmented Solow model (human capital is the augmenter in this neo-classical model); (2) a model using a Bayesian Averaging of Classical Estimates (BACE) approach following Sala-i-Martin, Doppelhofer and Miller (2004); and (3) Hanushek and Woessman's (2009) simple but persuasive model using a dataset of educational quality derived from a variety of international testing programmes. Other empirical work is also discussed, but in less depth, where such work is related in some way to the three models that form the main focus of the section.

As will be seen, the three models each represent important advances in the use of country-level data to prove or disprove the theoretical positions that have been put forward, and which were discussed in section 2. Two of the models can be said to be especially widely known, and hence they were selected for closer examination in this section. Mankiw, Romer and Weil (1992) stands out as a pioneering text and is often a point of reference (or the target of an antithesis). Hanushek and Woessman have succeeded in entering the discourse of education policymakers to an extent that is unusual for education economists (not just due to their 2009 text, but also other work by both authors). The Sala-i-Martin, Doppelhofer and Miller (2004) text was also selected because it seemed interesting from a methodological point of view. In fact, the BACE approach described in this text is used in Part II of the dissertation to examine whether the education variables typically used in growth models are in fact the best ones.

This section limits itself to a particular and relatively self-contained area of empirical analysis focussing on growth. To quote Lucas (1988: 3), the models dealt with in this section deal with 'the problem of accounting for the observed pattern, across countries and across time, in levels and rates of growth in per capita income'. Within that focus, the emphasis of the dissertation is largely on what the models can tell the policymaker about the relationship between education and growth. Cross-country growth models are undoubtedly the central empirical focus in textbooks on growth such as that by Aghion and Howitt (2009). Yet, as pointed out in the introduction in section 1, it would be incorrect to consider this type of empirical work as the only way of exploring causal factors behind growth. In fact, Solow (2001), writing almost half a century after his original 'Solow model' text, expresses some surprise and unease over the way in which cross-country growth modelling has come to dominate the empirical enquiry into the factors behind growth. So what are the alternatives? In section 1, the influence exerted by microeconomic rates of return models was discussed, as well as the interpretation problems associated with these models. Reference was also made to the limited and tentative contribution made by computable general equilibrium (CGE) models focussing on dynamics within one country. Aghion and Howitt (2009: 309-11), in a departure from their primary focus on cross-country empirical models, discuss a *within*-country analysis when they consider data from the fifty states of the United States. Of course this type of within-country analysis is analogous to the cross-country approach insofar as comparison occurs across geographical units. Fedderke and Luiz (2008) provide an original within-country analysis, focussing on South Africa, where time series data on human capital, economic and institutional factors are used to explore directions of causation. Data covering the period 1917 to 1997 are used. Findings are based on both bivariate and multivariate analysis. For the latter, cointegration techniques are used. Amongst other things, it is found that in the relationship between human capital variables, on the one hand, and growth, on the other, impact flows mostly from the former to the latter. Moreover, variables that proxy for the quality of human capital, such as the proportion of graduates in more mathematically inclined fields, have a greater impact on growth than more quantity focussed variables, such as school enrolment ratios.

Despite its popularity, cross-country growth modelling has attracted its share of strong criticism. Krueger and Lindahl (2001) identify two points of controversy relating to the cross-country growth literature. The controversies arise partly because the micro and the macro, or cross-country, evidence point in different directions. One controversy relates to the relative importance of the different levels of education in the growth process. Cross-country analyses have often supported the need to focus on strengthening secondary school and post-school education, whilst the microeconomic literature points strongly to promoting basic primary schooling. Secondly, the cross-country evidence has often supported a key element of endogenous growth theory, namely that if one knows the level of education in the population at the start of a period, then one does not require data on the growth in the level of education in subsequent years because any such growth is explained endogenously. The theory is not often explained in this manner, but this is perhaps an appropriate way of explaining it. The more common way of putting it is that endogenous growth theory sees the initial level of education as important for growth and subsequent growth in the level of education as unimportant. This view has, understandably, provoked frustration around the logic of endogenous growth theory as seen in the following extract from Breton (2011: 767):

> Most endogenous theories of growth suffer from an implicit conceptual problem. They assume that a nation's level of human capital affects its rate of growth over a subsequent period, but the length of the period is not defined. As the time period becomes longer, it becomes increasingly difficult to accept the theories' implicit assumption that increases in human capital during the period are not affecting the rate of growth.

Krueger and Lindahl (2001: 1130) conclude on the basis of their own data analysis that measurement error has been a hidden and important factor that has led endogenous growth proponents, such as Benhabib and Spiegel (1994), to believe initial level of education is somehow more important than educational growth. Once measurement errors are taken into account, cross-country models do attach independent explanatory power to growth in human capital. In contrast, the initial level of human capital is only important for the growth of those countries starting off from an exceptionally low level of economic development.

At a more fundamental level, Krueger and Lindahl (2001) dismiss the finding reached by some, notably Pritchett (2001) in an article titled 'Where has all the education gone?', that there is no significant association at all between educational attainment and growth. Such a finding, they argue, is simply the result of an insufficient sensitivity to data quality. Once data problems have been taken into account, an association that is economically important is virtually unavoidable.

Partly as a result of the selection of models for this section, these debates from the cross-country literature do not feature strongly in this section. Mankiw, Romer and

Weil (1992) subscribe to the neoclassical framework that predates endogenous growth theory, whilst the other two texts discussed below are not clearly or strongly aligned to endogenous growth theory.

The econometric questions in this section are partly dealt with by assessing the degree to which four frequently mentioned specification errors occur, or are successfully avoided. Following Gujarati (2003: 509) the four are:

- **Omission of a relevant variable** (or underfitting a model). If an explanatory variable that should be present is omitted in the growth regression the intercept and possibly the slope coefficients of the incomplete model will be biased. To avoid this error, the analyst should consider the underlying theory carefully and ensure that all variables that are important in the theoretical model also appear in the empirical model.

- **Inclusion of an irrelevant variable** (or overfitting a model). Including a variable that should be omitted within the growth regression may not influence the magnitudes of the intercept or the slope coefficients of the relevant variables. However, the likelihood that the intercept and relevant slope coefficients will be statistically significant is reduced. Again, the solution is to pay close attention to the theoretical model and not to include explanatory variables that do not directly explain the dependent variable, or that duplicate other explanatory variables.

- **Adopting the wrong functional form.** This error arises when variables are entered into the growth regression using an inappropriate form, for instance a linear form when the natural logarithm would be more appropriate. The recommended remedy here is to try out different forms and to examine results such as the R^2 statistics and t values. However, the analyst can also go a step further and apply standard tests such as the Durbin-Watson test, the Ramsey RESET test or Lagrange Multiplier (LM) test.

- **Errors of measurement.** Compared to the above three specification errors, the problems with one's equation that arise because there are measurement errors in one's variables, are particularly difficult to resolve. As explained by Gujarati (2003: 524), even randomly distributed measurement errors in one's explanatory variables bias the values of the coefficients on these variables, meaning one's equation is strictly not suitable for describing the relationships in one's model. Specifically, the coefficient values will be biased towards zero. This phenomenon is sometimes referred to as regression dilution or attenuation bias (Krueger and Lindahl, 2001: 1115). The problem is somewhat reduced if one's measurement errors are found only in the dependent variables, as long as the measurement errors are random. In this situation, the values of one's coefficients will not be biased, though the standard errors will be. To minimise biases caused by measurement error, the analyst should ideally have a thorough understanding of how the data were generated, of the likelihood of measurement errors in general, and of the probability that the errors are randomly distributed.

A fifth error mentioned by Gujarati (2003: 529), incorrect specification of the stochastic error term, is not dealt with here as it did not appear to be a relevant issue in the selected models. This error is said to occur in a situation where the usual form of the error term in the regression, as an added term, should in fact be replaced by a multiplicative form where the error term is multiplied by one of the explanatory variables.

4.2 The augmented Solow model of Mankiw, Romer and Weil

As explained in section 2.3 above, the intention in Mankiw, Romer and Weil (1992, abbreviated as MRW) was largely to improve the compatibility of Solow's (1956) model with a non-convergence scenario, where countries do not eventually converge to the same level of income per capita. As already explained, MRW achieve this through the introduction of a human capital variable to Solow's (1956) original ('un-augmented') model. Closely related to this first intention was a second intention, to adapt Solow's theory so that its implicit magnitudes approximate better the magnitudes found in the real world (MRW, 1992: 408). In terms of the model specification issues mentioned in the introduction to this section, MRW introduce a variable, namely human capital, that had been omitted from the model. They establish that the variable is indeed important as its inclusion results in estimates whose magnitudes are closer to those of the real world (details in this regard are discussed below).

MRW, arguably to a greater extent than the authors of the other two texts discussed in detail in section 4, base their empirical model on a careful consideration of the underlying theory. The key mathematical arguments leading to MRW's empirical model are reproduced here, with some simplifications and adaptations, partly to align the notation here with what was used in discussing MRW's theoretical model in section 2.

To begin with, we have the two determinations of physical capital (K) and human capital (H) in year t indicated in equations (39) and (40). Both equations follow the logic of earlier equation (4). To take human capital as the example, the amount of human capital is a positive function of the proportion of total national income during the previous period being invested in education, and is a negative function of the proportion of human capital that was lost during the previous period, through deaths and net emigration.

$$K_t = K_{t-1} + s_k Y_{t-1} - \delta K_{t-1} \tag{39}$$

$$H_t = H_{t-1} + s_h Y_{t-1} - \delta H_{t-1} \tag{40}$$

The above can be simplified by removing what is ultimately our dependent variable, total income Y, and making use of the exponents α and β from equation (6), which gives us the following:

$$K_t = K_{t-1} + \left(s_k^{1-\beta} s_h^{\beta} K_{t-1}\right)^{\alpha-\beta} - \delta K_{t-1} \tag{41}$$

$$H_t = H_{t-1} + \left(s_k^{\alpha} s_h^{1-\alpha} H_{t-1}\right)^{\alpha-\beta} - \delta H_{t-1} \tag{42}$$

The future steady state implied by the original Solow model, in which the amount of physical capital does not increase any more and investment in physical capital equals depreciation in physical capital, is captured in equation (43), which is based on equation (41). There is an equivalent equation for human capital.

$$\left(s_k^{1-\beta} s_h^{\beta} K_{t-1}\right)^{\alpha-\beta} = \delta K_{t-1} \tag{43}$$

$$\left(s_k^{\alpha} s_h^{1-\alpha} H_{t-1}\right)^{\alpha-\beta} = \delta H_{t-1} \tag{44}$$

Equation (43) can be restated as equation (45), and likewise equation (44), the human capital equation, can be restated as equation (46). The time subscripts are excluded here in the interests of simplicity.

$$K = \left(\frac{s_k^{1-\beta} s_h^{\beta}}{\delta}\right)^{\frac{1}{1-\alpha-\beta}} \tag{45}$$

$$H = \left(\frac{s_k^{\alpha} s_h^{1-\alpha}}{\delta}\right)^{\frac{1}{1-\alpha-\beta}} \tag{46}$$

Combining equation (46) and the original production function in equation (7) allows for the following equation which, as we shall see, lends itself to a log-linear regression form. Equation (47) would yield the same value *Y* as equation (7) if one assumes that a steady state of income prevailed and that therefore equations (45) and (46) held true.

$$Y = \frac{s_k^{\alpha/(1-\alpha)}}{\delta^{\alpha/(1-\alpha)}} \times H^{\beta/(1-\alpha)} \tag{47}$$

At this point we can formulate the following equation using natural logarithms where the only variables within the three slope coefficients are α and β (and α and β do not appear outside the slope coefficients). Equation (48) can clearly be used as a growth regression model if we have data on income *Y* and on the three explanatory variables: the physical capital savings rate (s_k), the capital depreciation rate (δ) and the amount of human capital (*H*).

$$\ln(Y) = \frac{\alpha}{1-\alpha}\ln(s_k) - \frac{\alpha}{1-\alpha}\ln(\delta) + \frac{\beta}{1-\alpha}\ln(H) \tag{48}$$

But MRW do not assume technology (*A*) and labour (*L*) from equation (7) to be 1. In fact, they assume that both have growth rates, referred to as *g* for *A* and *n* for *L.* Moreover, MRW aim to solve a growth regression where income per capita, or *Y* over *L,* is the dependent variable, given that this is the logical measure of a country's level of economic development. Equation (49) restates the regression model in such a way that these requirements are taken into account.

$$\ln\left(\frac{Y_t}{L_t}\right) = \ln(A_{t=0}) + gt + \frac{\alpha}{1-\alpha}\ln(s_k) - \frac{\alpha}{1-\alpha}\ln(n+g+\delta) + \frac{\beta}{1-\alpha}\ln(H) \tag{49}$$

It can be shown that equation (49) is true if equation (7) and the following expanded version of equation (46) are true:

$$H = \left(\frac{s_k^{\alpha} s_h^{1-\alpha}}{n + g + \delta} \right)^{\frac{1}{1-\alpha-\beta}} \qquad (50)$$

Equation (49) can be simplified to the following form:

$$\ln Y_i = B_0 + B_1 \ln s_i + B_2 \ln(n + g + \delta) + B_3 \ln H_i + u_i \qquad (51)$$

Here s_i is what has been referred to earlier as s_k for country *i.* The form in equation (51) is thus a log-linear version of the classical linear regression model. The first two terms on the right-hand side of equation (49) can be collapsed into one term, the natural log of *A* in year *t*. This is thus what B_0 represents. MRW assume that time *t* in equation (49) is a point in time when all countries have reached their steady state of per capita income and hence *A* in year *t* would carry the same value for all countries. They thus assume that technology possessed by one country will very easily be possessed by others too.

In selecting data, MRW use for the dependent variable the natural logarithm of real GDP over working age population in 1985. The explanatory variable *s* is represented by the average gross capital formation, both public and private, over GDP, for the years 1960 to 1985. For *n* MRW use average annual growth in the population of working age, or aged 15 to 64, for the years 1960 to 1985. For *g* they simply use the value 0.02 for all countries based on trends in the United States (MRW are treating *A* as a resource freely available to all countries). Similarly, they use a uniform value for capital depreciation δ of 0.03 for all countries, again based on US data. It is argued that these two variables are unlikely to vary substantially across countries. Finally, *H* is represented by the number of secondary school pupils divided by the population of working age. For this, the average across the entire growth period, 1960 to 1985, is used. MRW admit that this is an imperfect measure of human capital, but argue that it might be the best that is available. Finally, *u* is the error term. What all this means is that only *s*, *H* and *n* are left as country-specific predictors of aggregate income.

Below, the function in equation (51) is tested using more recent data than that used by MRW for *Y*, *s, H* and *n*. For *g* and δ the original MRW values are used. Data from the years 1980 to 2005 are used, in other words from a period that is 20 years later than that of MRW. Whether the function has a general applicability beyond the historical period considered by MRW is thus tested. Moreover, the effects of a few simple adjustments to the data are tested. Further on the effect of using a completely different measure of human capital is tested. The following table describes the data used. The original variables used by MRW are marked with **. The remaining rows indicate variations on the original variables.

Table 5: Variables for MRW model

Variable	*Source*	*N*	*Mean*	*Min.*	*Max.*
GDP (in real PPP USD of 2006) over population aged 15 to 64, average 2002-2004**	Heston, Summers and Aten, 2006.	172	14754	811	73796
GDP per capita (in real PPP USD of 2006), average 2002-2004	As above.	183	9688	399	49536
Gross fixed capital formation (public and private) over GDP, average 1980-2005**	Online data-querying facility of United Nations Statistics Division (National accounts estimates of main aggregate), consulted 2009.	200	23.0	8.7	87.6
Annual increase in working age population (age 15 to 64), average 1980-2005[42]**	UN Population Division, 2009.	174	2.15	-0.30	7.53
Percentage of working age population in secondary schooling, average 1980-2005**	Online data-querying facility of UNESCO: UIS, consulted 2009.	172	8.98	0.88	21.70
Percentage of working age population in secondary schooling, average 1980-1990	As above	161	8.34	0.65	21.70
Gross enrolment ratio in secondary schooling, average 1980-1990	As above	174	54.7	3.6	159.0
Note: All population values are online data-querying facility of United Nations Statistics Division (World population prospects: The 2008 revision), consulted 2009.					

Altogether 211 countries have values for at least one variable in the above list, 170 have values for all the variables marked **. MRW excluded eight countries whose incomes were derived mainly from oil so that countries with economies that would clearly not be explained by equation (51) would not contaminate the results. In the analysis that follows, any country with oil exports worth more than a third of total exports in 2005 according to UN data was excluded from the analysis. Two further exclusions employed by MRW were of countries with fewer than one million inhabitants and communist countries. The former were considered likely to display abnormal growth patterns, presumably due to factors such as extreme trade openness or special relationships with larger metropolitan powers, and the latter presumably due to abnormal constraints on the returns to human capital resulting from high levels of government control over profit-maximising behaviour by individuals. When countries with fewer than one million inhabitants in 2000 were excluded from the dataset used here, the 170 countries referred to earlier were reduced to 142. Applying the oil exportation criterion further reduced the sample to 129 countries. Finally, excluding ex-communist countries (there were no currently communist countries in the sample) reduced the sample to 108 observations. The MRW sample size after similar exclusions was 98.

MRW furthermore excluded countries if the authors of the Penn dataset used by MRW considered the income data to be unreliable. As the 2006 version of this dataset (Heston, Summers and Aton, 2006) used here did not include indicators of poor data quality, these exclusions could not be repeated, and depending on improvements in the Penn data, they may not be necessary.

[42] Values, which were calculated from population by five year bins, could have been obtained from a greater number of countries from this data source. However, values were only calculated when the availability of other values for the country suggested that the country would be included.

The following table provides results obtained by MRW (1992: 420) and results from a model that uses the ** variables from the previous table and thus replicates the MRW model as faithfully as possible using later data (see Model A). Using data from a later time period seems to make little difference to the original MRW findings. All three explanatory variables remain highly statistically significant and carry the same sign as in the original model. The overall fit of the model as measured by the adjusted R^2 remains high.

Table 6: Regression outputs – replication of MRW

Variable	*Original MRW*	*Model A*	*Model B*	*Model C*
Constant	6.89***	8.37***	8.01***	4.65***
	(5.89)	(3.73)	(8.42)	(4.59)
ln(*s*)	0.69***	1.02***	1.12***	1.09***
	(5.31)	(3.73)	(4.25)	(4.17)
$\ln(n + g + \delta)$	-1.73***	-0.77***	-0.72***	-0.50***
	(-4.22)	(-7.74)	(-7.04)	(-4.36)
ln(*H*)	0.66***	0.87***	0.77***	0.89***
	(9.43)	(7.40)	(7.75)	(9.54)
N	98	108	105	106
R^2		0.712	0.729	0.771
Adjusted R^2	0.78	0.704	0.721	0.764

*Note: Dependent variable is GDP per adult in 2002. *** indicates that the estimate is significant at the 1% level of significance. Values in brackets are t statistics.*

Model B deviates from the MRW approach insofar as *H* is based on data only for the earlier 1980 to 1990 period (in Model A, the average across the entire growth period 1980 to 2005 was used). One might expect this to result in a better fit of the data to the theoretical model given the time lag between higher secondary school enrolments and better human capital in the labour market. The results from this model are hardly different from those of Model A, though a slightly better fit as measured in the adjusted R^2 is seen. Taking Model B and alternately re-inserting small, oil-exporting and ex-communist countries in three separate models makes no substantial difference to the levels of significance (all three variables remain significant at the 1% level), though the overall fit of the model declines substantially with adjusted R^2 dropping to 0.56, 0.65 and 0.56 respectively. This suggests the exclusions are indeed necessary, in particular with respect to small and ex-communist countries.

Model C converts two of the MRW variables to a more traditional format. GDP per capita, instead of GDP over working population, is used. Moreover, what UNESCO would call the gross enrolment ratio (GER) for secondary schooling, which is secondary enrolments over the population one would assume is in secondary school if everyone attended (and there was no grade repetition), is used instead of secondary enrolments divided by working age population. MRW chose their forms on the basis of theoretical considerations. As the results for Model C show, the more traditional forms substantially improve the overall fit of the model as reflected in the adjusted R^2. This may be the result of errors of measurement problems in the age-specific population statistics required for the two denominators in the MRW model, rather than any problems with MRW's theory.

The results in Table 6 do not answer the question of what the actual values of α and β are in equation (49). Using simple algebra and equation (49) to calculate the two values turns out not to be a satisfactory approach as ambiguous results are obtained. MRW resolve the problem through the use of a restricted least squares form that is permitted by the fact that B_1 and B_2 in equation (51) are in theory equal to each other as they are based on the same factor in equation (49). The restricted form used by MRW (1992: 420) is as follows:

$$\ln Y_i = B_0 + B_1(\ln s_i - \ln(n + g + \delta)) + B_2(\ln H_i - \ln(n + g + \delta)) + u_i \qquad (52)$$

The resultant value for α is 0.31 and for β it is 0.28 in MRW. This, they argue, underlines how a relevant variable, H, has been omitted in the textbook Solow model. If human capital did not matter, β would be zero. Omitting the human capital variable from the textbook Solow model means that this model over-estimates the impact of saving s on growth. Without human capital and without β the exponent α in the production function becomes 0.60, or twice as large as it should be. When the same restricted regression used by MRW was applied to the later data, the importance of human capital relative to physical capital represented by s is emphasised even more strongly than in the original MRW analysis. Specifically, the model without human capital yields α equal to 0.54, or a value similar to the corresponding value in the original MRW, whilst inserting human capital reduces α to 0.09, or one-sixth of what it was in the textbook Solow model. β, or the human capital exponent in equation (7), equals 0.41 in the restricted model using the later data, against 0.28 in the original MRW analysis. It thus seems as if a historical rise of human capital as a determinant of economic success is borne out by the evidence presented here.

What about the question of convergence amongst countries to similar ultimate levels of income? As has been pointed out in section 2.2, the un-augmented textbook Solow model does permit non-convergence, or different steady states of income for different countries, if initial capital endowments and the rate of saving are different. MRW use the regression in equation (53), which is an adaptation of equation (51), to investigate how well their augmented model predicts the magnitude of income growth that different countries have experienced. The variable y, or log of income per working age person at the start of the period in question (1960 in MRW, 1980 in the new analysis presented here), is inserted as an explanatory variable, whilst the dependent variable becomes the difference between the log of income at the end and start of the period. Critically, if B_1 is negative and significant, then this would suggest that those starting off with a lower income gain more in absolute terms (and, by implication, experience higher annual growth rates).

$$(\ln Y_i - \ln y_i) = B_0 + B_1 \ln y_i + B_2 \ln s_i + B_3 \ln(n + g + \delta) + B_4 \ln H_i + u_i \qquad (53)$$

The next table displays the results for the above regression model from the original MRW analysis and for adjusted versions of models A to C from Table 6.

Table 7: Regression outputs – tests of convergence

Variable	*Original MRW*	*Model A*	*Model B*	*Model C*
Constant	3.04***	-0.44	-0.51	-1.48*
	(3.66)	(-0.54)	(-0.62)	(-1.91)
ln(*y*)	-0.29***	-0.11*	-0.10*	-0.08
	(-4.66)	(-1.75)	(-1.68)	(-1.17)
ln(*s*)	0.52***	0.89***	0.91***	0.92***
	(6.02)	(5.95)	(6.09)	(5.96)
ln(*n* + *g* + *δ*)	-0.51*	-0.23***	-0.23***	-0.14*
	(-1.75)	(-3.25)	(-3.18)	(-1.84)
ln(*H*)	0.23***	0.06	0.05	0.11
	(3.88)	(0.75)	(0.66)	(1.34)
N	98	101	100	101
R^2		0.371	0.373	0.377
Adjusted R^2	0.46	0.345	0.346	0.351

*Note: Dependent variable is difference in GDP per adult between 2002 and 1980. *** indicates that the estimate is significant at the 1% level of significance, * at the 10% level. Values in brackets are t statistics.*

The key finding from MRW of a negative and significant value for B_1 is found in two of the models using the later data. The most notable difference between the original results and those using the recent data is that the human capital variable appears non-significant in the latter across all three models. This is in contrast to the models predicting just level of income in Table 6, where the human capital variable was consistently highly significant. As will be seen, this is largely due to errors of measurement inherent in the use of school enrolments as a proxy for actual development of human capital. In fact, one would expect this problem to have got worse in the later period as inequalities with respect to secondary schools enrolments diminished whilst gaps in the quality of education remained large.

What would one find if the measures of what pupils learn in schools discussed in section 3.3 are used in the augmented Solow model? Results appear in Table 8. The two models described are variations of Model C from Table 6. The difference is that measures of the quality of human capital are used here. Model C1 uses values calculated by Hanushek and Woessman (2009), specifically their values for the quality of schooling at the lower secondary level (these yielded the best fit of the different educational quality variables calculated by these authors). Model C2 uses the newly calculated values for H described in section 3.3 (they are divided by 100 to give a similar metric to that used in Model C1). Model C (and not A or B) was selected for Table 8 as this model yielded the best fit in terms of adjusted R^2. Both models in Table 8 represent a worse fit, in terms of adjusted R^2, than any of the models in Table 6. This pattern persists even if one runs Model C from Table 6 using only the countries used in the models in Table 8. Most striking, however, is the low level of significance of s in Table 8. This is largely due to multicollinearity between the new human capital variables and investment in physical capital s, a problem which did not occur when secondary enrolment was used for H. To illustrate, the correlation coefficient between the Hanushek and Woessman (2009) lower secondary quality variable and s is 0.31, against a correlation between s and secondary enrolments (H in Model A of Table 6) of just 0.01. These two correlation coefficients were calculated using the same set of countries, as many as the data permitted. This seems telling. A plausible explanation is perhaps that a more learned population is more likely to see the advantages of investment in physical capital and hence saves more towards this end.

Table 8: Regression outputs – MRW with quality of human capital

Variable	*Model C1*	*Model C2*
Constant	5.31***	4.52***
	(3.48)	(2.91)
ln(*s*)	0.67	0.69*
	(1.48)	(1.78)
ln($n + g + \delta$)	-0.40***	-0.41***
	(-2.91)	(-3.13)
ln(*H*)	2.67***	3.13***
	(3.96)	(5.63)
N	52	68
R^2	0.604	0.619
Adjusted R^2	0.579	0.601

*Note: Dependent variable is difference in GDP per adult between 2002 and 1980. *** indicates that the estimate is significant at the 1% level of significance, * at the 10% level. Values in brackets are t statistics.*

Where using a measure of the quality of human capital makes a large difference is in predicting actual economic growth within individual countries, using the tests of convergence models put forward in Table 7. Table 9 provides variations of Model A from Table 7 using the two human capital variables used for Table 8. Model fit as seen in the adjusted R^2 is better than that of the original MRW model (0.53 compared to 0.46). Instead of the non-significant human capital variables using recent data seen in Table 7, here a highly significant human capital effect is seen. Moreover, the explanatory variable *y* appears highly significant here, confirming that conditional convergence between countries is a reality, as indicated by the original MRW model of Table 7.

Table 9: Regression outputs – tests of convergence with educational quality data

Variable	*Model A1 (H&W HC)*	*Model A2 (new HC)*
Constant	-0.99	-1.03
	(-1.23)	(-1.41)
ln(*y*)	-0.20***	-0.16***
	(3.53)	(-3.40)
ln(*s*)	0.83***	0.64***
	(4.40)	(4.33)
ln($n + g + \delta$)	-0.12**	-0.10*
	(-2.09)	(-1.81)
ln(*H*)	0.84***	0.97***
	(2.83)	(4.21)
N	49	63
R^2	0.566	0.494
Adjusted R^2	0.526	0.459

*Note: Dependent variable is average annual growth 1960-2004. *** indicates that the estimate is significant at the 1% level of significance, ** at the 5% level, and * at the 10% level. Values in brackets are t statistics.*

This section has offered a few permutations of a famous and pioneering growth regression. It was seen that the findings of the original 1992 implementation of the model by MRW are replicated if later data are used. More interestingly, the basic findings are replicated if a measure of school test results, reflecting the situation across the world in the mid-1990s (see section 3.3), is used to represent human capital. The correlation between investment in physical capital and the test score measure is considerably higher than the correlation between investment in physical capital and MRW's original human capital variable, which reflected the level of enrolment at the secondary level. It was speculated that this could be because actual human capital in the form of knowledge, as opposed to just years of schooling, is a better predictor of people's understanding of the importance of saving. Though there are likely problems with the use of a measure of human capital that is from such a recent period (this is discussed further in section 4.4 below), the findings presented above add to the evidence that statistics from testing programmes are valuable if one wants to model growth and that within the growth process, what people actually know is a vital element of human capital.

4.3 The BACE model with iterative variable selection

The analysis that principally guides this sub-section is that of Sala-i-Martin, Doppelhofer and Miller (2004), titled 'Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach'. This text is henceforth referred to as SDM. What also informs this section is two files accompanying SDM, one with code in GAUSS, a matrix programming language, and the second a cross-country dataset covering 139 countries[43].

Of all the empirical models considered in section 4, that of SDM is arguably least rooted in the economic theory presented in section 2. In a sense this is a natural outcome of their intention to let the data speak for themselves (SDM, 2004: 813).

SDM concentrate on two of the four specification errors referred to at the start of section 4, namely the omission of a relevant variable and the inclusion of an irrelevant one. Informing their approach is their criticism that cross-country growth analysts tend to select a very limited set of explanatory variables from the larger set of all possible variables in a manner that is 'artistic' (SDM, 2004: 813) rather than scientific. SDM's approach is to study jointly 67 variables suspected of playing a role in economic growth within a rational methodology. A problem inherent within cross-country regression analysis is that relative to the number of possible explanatory variables, the number of observations, or countries, is too low to allow for meaningful inclusion of all variables simultaneously within one regression model. SDM's solution is essentially to run many regressions (in fact, tens of millions) using different combinations of explanatory variables, with growth as the dependent variable, and then drawing conclusions about the optimal model and optimal variables based on the entire exercise. SDM thus address an important econometric problem related to small samples and not really addressed by the other empirical analyses discussed in this section.

SDM's approach involves departing from pure classical statistics and using a method that involves some classical statistics and some Bayesian statistics (these two being the principal branches of statistics). Some discussion of what Bayesian statistics entails is provided as an introduction to SDM's method. Whilst classical statistics focuses on testing hypotheses, such as whether a parameter equals zero, Bayesian

[43] Available at http://www.nhh.no/Default.aspx?ID=3075.

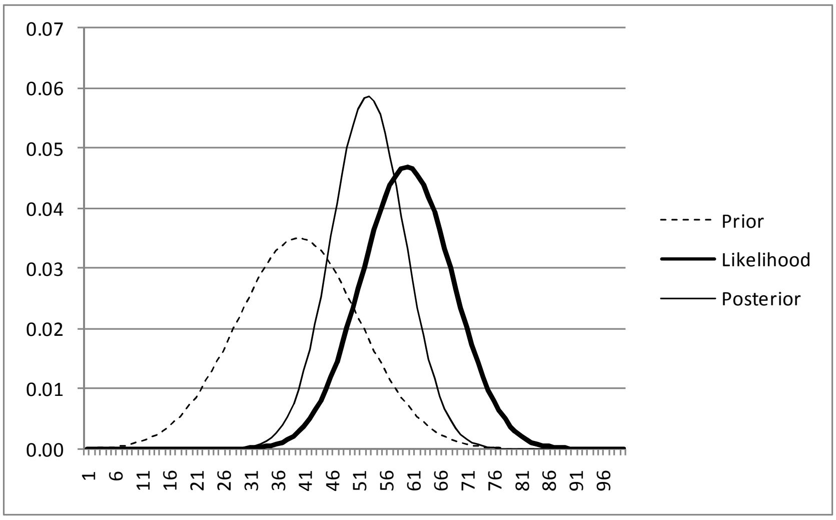
statistics focuses on producing the likely distribution of the unknown parameter, called the posterior distribution. Bayesian statistics uses a combination of the data and the assumptions or judgements of the analyst, where the latter is known as the prior distribution, to determine the posterior distribution. An advantage of Bayesian statistics over the classical approach is said to be that it makes assumptions more explicit. The counter-argument against Bayesian statistics is often that this approach is unscientific in over-emphasising expert opinion, relative to the data. Kennedy (2008: 213) refers to the 'violent controversy' surrounding which of the two branches of statistics is more justified. Kennedy (2008) moreover argues that it is not so much questions of correctness that have allowed classical statistics to remain dominant, but rather questions relating to the computational complexity of Bayesian statistics and the difficulty of explaining the approach to, for instance, policymakers.

Somewhat more formally, Bayesian statistics uses as its point of departure the Bayes theorem captured in equation (54) below. The left-hand side refers to the posterior distribution, or the distribution of the parameter β conditional on the data y, or after having taken account of the data. This is calculated by multiplying the prior distribution $g(\beta)$, or the distribution of β before seeing the data, by what we can understand as the actual data or what Kennedy (2008: 214, 226) refers to as the likelihood function or 'the "conditional" density of the data, conditional on the unknown parameters'. The denominator on the right-hand side is a 'normalization factor'.

$$g(\beta \mid y) = \frac{f(y \mid \beta)g(\beta)}{f(y)} \tag{54}$$

Following Kennedy (2008: 215) it is possible to demonstrate, for instance in Excel, the basic manner in which a prior distribution multiplied by a likelihood function divided by a constant results in the posterior distribution, which we can understand as the average between the other two distributions. The result of such a demonstration appears below.

Figure 16: The derivation of the Bayesian posterior distribution



Note: The horizontal axis refers to percentiles. The vertical axis refers to the Y of the equation for a normal distribution $Y = \frac{1}{s\sqrt{2\pi}} e^{\left(-(x-\bar{x})^2\right)/\left(2s^2\right)}$ *(Blalock, 1979: 93). For the prior distribution a mean of 40 and standard deviation (s) of 8 was used and for the likelihood function the values were 60 and 6. The normalization factor was set to ensure that the sum of Y over the percentiles for the posterior distribution equalled that found in the prior or likelihood curves.*

A problem that SDM have with using just the Bayesian approach in solving their problem is that this approach allows for too much arbitrariness. They thus resort to a combination of classical and Bayesian statistics where just one 'prior hyper-estimation' (SDM, 2004: 815) is used, this being an initial expert assumption on the likely size, in terms of number of explanatory variables, of the optimal regression model. This variable will be called *k* in the discussion that follows. SDM call their method Bayesian averaging of classical estimates, or BACE.

The 67 possible explanatory variables in SDM's dataset cover factors relating to the economy (such as trade openness), politics (such as degree of political liberties), geography (such as closeness to the equator) and culture (such as predominant religions). With respect to education, there are two variables, one being the primary enrolment ratio in 1960 and the other the tertiary education enrolment ratio in 1960. There are 13 0-1 dummy variables. The dependent variable is average real GDP per capita growth between 1960 to 1996. When observations with missing values are removed, the original 139 observations are reduced to 88 observations. All of SDM's analysis deals only with these 88 observations, 24 of which refer to high income countries.

The end result of SDM's method is a ranking of all 67 explanatory variable according to their ability to explain growth. A cut-off is drawn after the first 18 variables, these being considered the only ones with a significant relationship to growth, but where 'significant' is understood in a more general and non-classical manner (SDM, 2004: 823). Table 10 at the end of this section lists the 18 variables (see the first panel). The abbreviated variable names are explained in Table 12. The three top ranked explanatory variables are the East Asian dummy variable, primary enrolments in 1960 and investment price levels in 1960. The tertiary education variable mentioned earlier is not amongst the 18 significant variables. The values used to sort the 67 variables are the posterior inclusion probability values, which range from 0.823, in the case of the East Asian dummy, to 0.015.

Much of the discussion in this sub-section relates to how the posterior inclusion probability is calculated. This is done mainly in terms of the mechanics of the approach expressed in the GAUSS code accompanying SDM and less in terms of Bayesian theory (most of SDM's explanation of the approach is in terms of the latter). The GAUSS code was translated to Mata, a matrix programming language embedded in Stata, partly to examine whether SDM's implementation of BACE was in any way sensitive to the programming language used (this was not found to be the case to any noteworthy degree). The Mata code appears in Appendix B. Apart from the posterior inclusion probability, six other statistics are reported by SDM in their final results (SDM, 2004: 824). Of these, just two key ones are dealt with in this discussion and the Mata code, namely the posterior mean conditional on inclusion and the posterior standard deviation conditional on inclusion.

BACE involves the iterative running of many regression analyses of the data. Around 50 million such analyses must be run in the case of the SDM dataset. Every analysis is a simple OLS regression, but with a different combination of explanatory variables used each time. It is possible for the same combination of variables to be repeated, though the probability of this is extremely small, given the number of possible combinations (discussed below). The choice of variables is at first completely random, meaning each variable has an equal probability of entering the model, but as the process proceeds more non-random selection elements come to play. How this occurs is a key aspect of SDM's method. A model could contain between 1 and all 67 explanatory variables, meaning the total possible combinations is 2^{67} or around 1.5 by 10^{20}. So even 50 million combinations cover only a very small fraction of all possible combinations. Initially the number of explanatory variables clusters around *k*, with *k* being both the mode and median model size. The value SDM use for *k*, or the 'prior hyper-estimation', is 7. This is based on the analysts' experience of cross-country growth modelling. How a departure from the initial clustering around *k* equals 7 occurs, is a further key element of the BACE method. A replication of the entire SDM analysis took around seven hours on a fast laptop in 2011 using Mata within Stata 9.

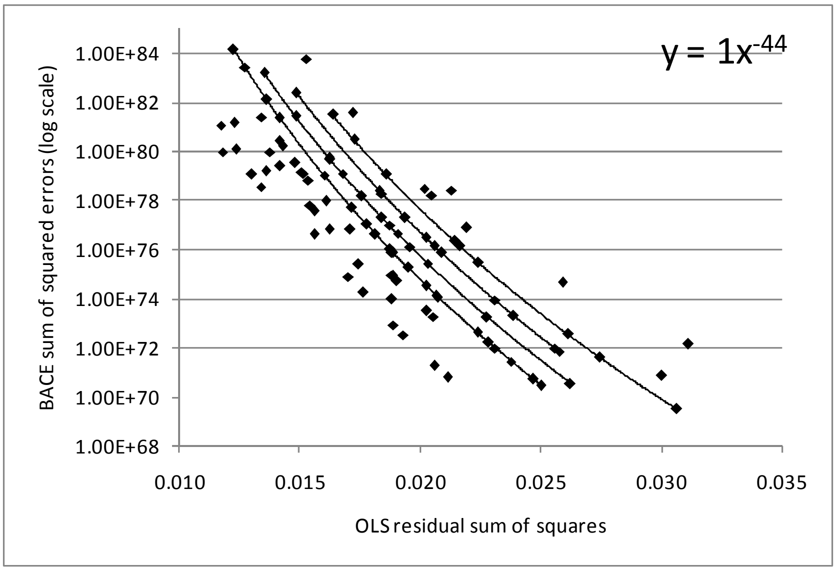
The final posterior inclusion probability statistic (see vector postprob in Mata) attached to each explanatory variable at the end of the BACE process is calculated as follows, where *v* is the variable in question.

$$p_v = \frac{\sum e^*}{\sum e} \tag{55}$$

Each value *e* is specific to one regression analysis and is referred to in SDM's programming code as the 'sum of squared errors'. Normally the sum of squared errors would be synonymous with what is probably most commonly referred to as the residual sum of squares (RSS) in an OLS regression. The higher the RSS relative to the total sum of squares (TSS) in a regression, the lower the R^2 value and the explanatory power of the model. As we shall see, SDM's sum of squared errors is closely related to the common RSS, but there are some key differences. The denominator in equation (55) is the sum of the various values *e* obtained across the 50 million or so regressions. The numerator Σe^* is the sum of the *e* values of just those models where variable *v* was included (in Mata e^* is sumess whilst *e* is sess). The implication is thus that p_v increases in value when either of the following happens: (a) variable *v* is included in a greater proportion of the regression models (if variable *v* were included every time, then p_v would equal 1.0); (b) The value *e* in the models where variable *v* is present is larger relative to the value *e* in the models where *v* is not present. How these effects come about is explained below.

To begin with, effect (b) is explained, as initially when the code is run it is just this effect, and not effect (a), that plays a role. The situation changes after the first 100,000 loops (or regressions). Importantly, SDM's sum of squared errors is negatively correlated to RSS. The following graph illustrates the relationship between RSS and SDM's sum of squared errors during the initial 100 loops. Each point represents a regression analysis. As explained above, the most common model size is 7 explanatory variables (model size is indicated by the curves in the graph).

Figure 17: The BACE sum of squared errors



Apart from the negative correlation between the variables on the two axes, what stands out is the very high values of the BACE sum of squared errors. They are in the range of around 10^{68} to 10^{84}. Returning to equation (55), the implication is that variables appearing in models with relatively low RSS values would be those variables with the best, or highest, posterior inclusion probability values. This makes intuitive sense as models with higher R^2 values are considered to carry greater explanatory power. The next two equations explain how RSS (denoted by *r* below) and *e* are related.

$$r = \sum_{i=1}^{n} (u_i)^2 \tag{56}$$

$$e = \frac{r^{\frac{-n}{2}}}{n^{\frac{m-k}{2}}} \tag{57}$$

Equation (56) is simply a reminder of how *r* is obtained. It is the sum of the squared errors across the *n* observations following the regression analysis. Equation (57) represents a calculation referred to in the notes in the code as 'ESS corrected by DFs a la Ron' (Ron being almost certainly a reference to one of the three authors, Ronald Miller). SDM's sum of squared errors is referred to as 'ESS' in the GAUSS code but as 'SSE' in the journal article. From the negative exponent in the numerator in equation (57) it is clear that the BACE sum of squared errors increases when *r* decreases. The denominator on the right-hand side of equation (57) varies only according to *m*, the number of explanatory variables selected for the current model. The number of observations *n*, which is 88 countries in SDM's dataset, and *k*, which SDM have set at 7, are constant. The variation in *m* results in the denominator taking on very small positive numbers for a small number of explanatory variables, for instance 0.0012 in the case of four variables, 1.0 for seven variables, and very large positive numbers in the case of a high number of explanatory variables, for instance 826 in the case of ten variables. The result is that for any given value of *r*, having fewer explanatory variables increases *e*. This is clearly visible in Figure 17. It is also intuitive. One would expect a model with two explanatory variables which has the same RSS as a model with four explanatory variables to be a better model given that generally RSS decreases the more variables you have (see SDM, 2004: 817). This is the one important dynamic within the relationship between *r* and *e*. The other is that the power function with a negative exponent results in a situation where there is an increasing marginal gain in the size of *e* as RSS decreases.

The focus now shifts to what occurs beyond the first 100,000 loops. Up to loop 100,000 the probability of being selected for the regression is the same for all variables. This probability is 0.104, or 7 over 67, giving on average of 7 selected variables per regression analysis. In loop 100,000, and thereafter after every 100,000 loops, something referred to in the code as the 'adjustment of sampling probability' occurs. What this means is that the probability of being selected (captured in the Mata vector `iprob`) is adjusted. The more a variable is associated with more predictive models, the higher its probability of being selected. However, no variable's probability of being selected is allowed to fall below a minimum of 0.100. By the end of the process most variables in fact carry a probability of 0.100 and a few carry high probabilities, the highest being around 0.8. This explains how effect (a), referred to above, becomes possible. The modal (most common) model size is 7 during the first 100,000 loops, 9 over loops 100,001 to 500,000, 10 for loops 500,001 to 1,000,000 and then 10 again for 49,500,001 to 50,000,000. There is thus a systematic increase in the size of the models. However, as explained below, when one calculates a weighted mean of the model size a somewhat different picture emerges.

Figure 18: The progression of model size and the BACE SSE

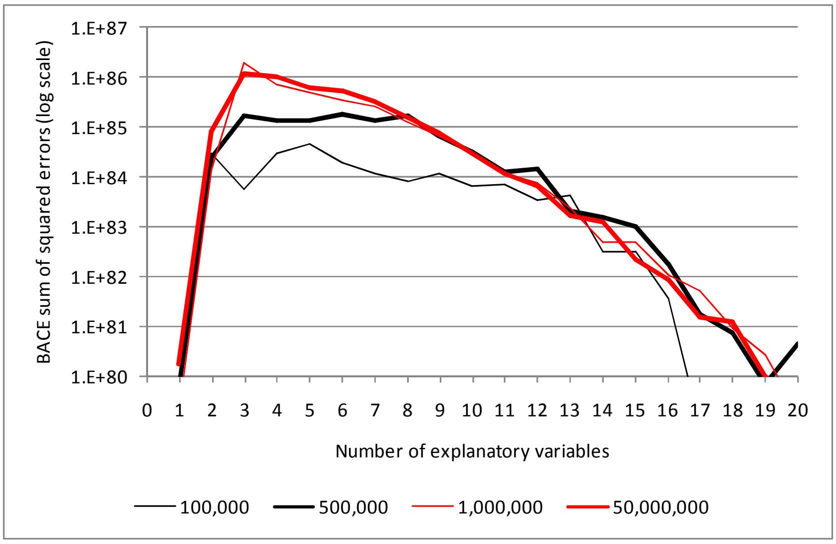


Figure 18 illustrates another trend across the four points in the process. Though the trend is towards more models with more than 7 explanatory variables, there is also a trend towards a maximum value of *e* associated with a model of size 3. This is to be expected given partly the fact that the selection probabilities for more predictive variables are rising and that equation (57) is designed to assign especially high values of *e* to smaller models.

Equation (58) below explains a weighting of *e* that takes effect in every loop after loop 100,000. Here *e* from equation (57) is replaced by ε to reflect the fact that after the first 100,000 loops, but not before then, there is a difference between what appears as *e* and ε below.

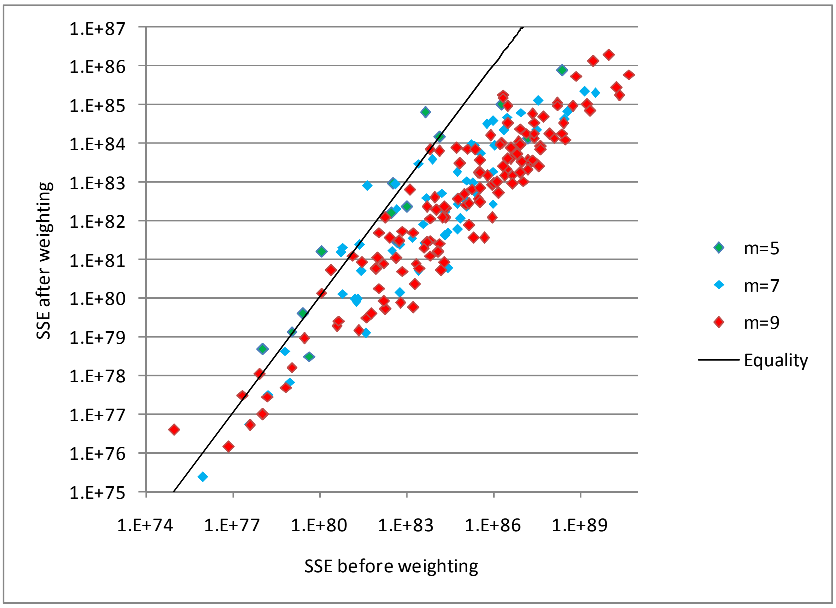
$$e = \varepsilon \times \left(\left(\frac{k}{q} \right)^{m} \times \left(1 - \frac{k}{q} \right)^{q-m} \right) \Big/ s \qquad (58)$$

The numerator of the weight in equation (58) includes q, the total possible number of explanatory variables (always 67), the prior assumed model size k (always 7) and the actual number of variables selected for the current model, m. The natural log of the numerator is in effect a negative linear function of m, so the greater the number of explanatory variables, the smaller the weight. The denominator s is calculated as follows:

$$s = \prod_{v=1}^{m} p_v \qquad (59)$$

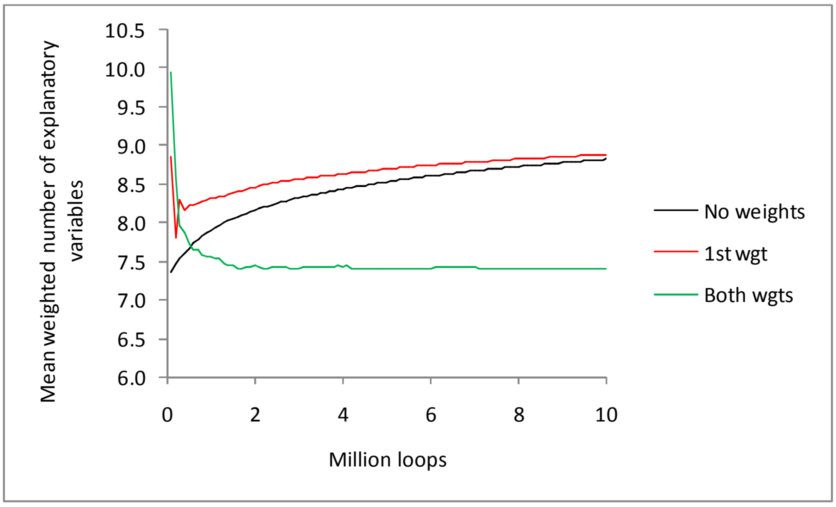
In other words, s is the product over all variables in the selected model of p_v, where p_v is the probability that this variable would be chosen. Models with 'weaker' variables thus have a lower value of s and hence a greater weight. During the first 100,000 loops, the weight is 1.0 because the numerator and denominator are equal. The following graph illustrates the effect of the weight later in the process. Specifically, a random selection of regression analyses from the 1,000 loops preceding the 3,000,000 mark is illustrated. The general effect of the weight is to reduce the value of *ε,* in particular for higher values of *ε* and where the number of selected variables *m* is higher.

Figure 19: The weighting of the BACE SSE



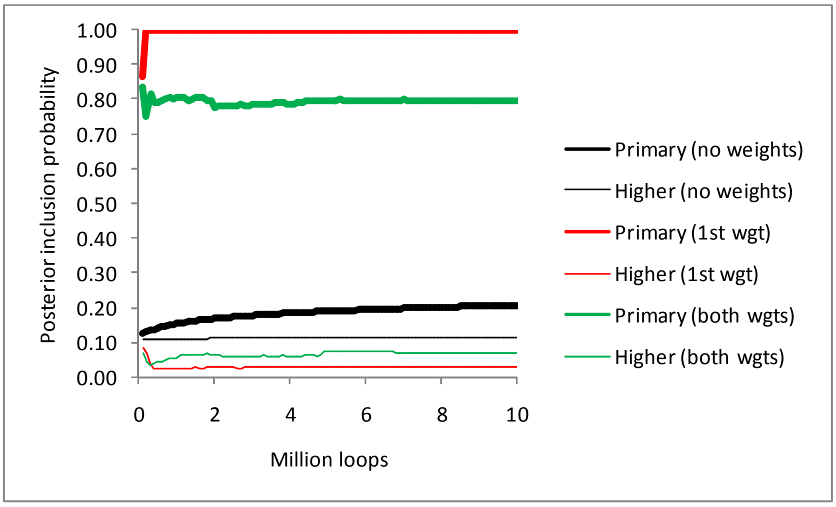
The next two graphs sum up the effects of the two transformations on the original *r* represented by equations (57) and (58). The graph reflects points for every 100,000 loops during the first 10 million loops, with each point reflecting the average over the preceding 100,000 loops. The average is the weighted average number of explanatory variables. The exercise involved creating two approaches, not used by SDM, the ‘No weights’ and ‘1st wgt’ scenarios, for demonstration purposes. For the ‘No weights’ curve, the inverse of *r* was used as the weight, in keeping with the idea that holding the number of explanatory variables constant, a greater *r* represents a less predictive model. Introducing just the weight implied by equation (57) produces, like the ‘No weights’ scenario, average model sizes that are well above the prior assumption of 7. It is only with the introduction of the additional weight from equation (58), in other words by following the full weighting used in BACE, that an average model size close to 7 is maintained for most loops. Figure 20 thus illustrates the justification for the calculation of *e* in BACE.

Figure 20: Effect of weights on mean model size



The three scenarios can be illustrated with respect to the performance of the posterior inclusion probability of variable *v*, or p_v. In the following graph the progression of the p_v values for the two education variables, the primary and tertiary education enrolment ratios, is illustrated. It is only when both weights are applied that p_v values reflecting realistic differences between the inclusion probabilities of the two variables emerge.

Figure 21: Effect of weights on posterior inclusion probability



The estimated size of the slope coefficient for each explanatory variable calculated in BACE is referred to as the 'posterior mean conditional on inclusion'. The method for obtaining this is relatively straightforward. The mean across the coefficients in the various models where variable *v* was present is used, with the mean being weighted by *e* from equation (58). It is in fact an analysis of the slope coefficients which is used to determine when the BACE processing should end. Basically, when the standardised beta versions of the slope coefficients have become sufficiently stable across loops, convergence is said to have been reached and the processing stops.

The 'posterior standard deviation conditional on inclusion' is calculated in a similar manner. Again, for each variable the mean, weighted by *e*, of the standard deviations of slope coefficients found across the regression analyses is used. The BACE slope coefficients and standard deviations are close to what would be obtained if one ran a once-off OLS regression analysis of the seven variables with the highest posterior inclusion probabilities.

The above has been a discussion of the key mechanics of the BACE method. Returning to the explanations by SDM in terms of Bayesian theory (or departures from this), a key matter is why *e* should be an appropriate measure of the 'strength' of a particular model. Put differently, what is the justification for equations (57) and (58)? Equation (57) is said to approximate a 'degrees of freedom correction' employed by Gideon Schwarz to avoid a situation where otherwise the 'posterior mean model size ... will always be bigger than the prior model size, irrespective of the data' (SDM, 2004: 817). This is something that was seen in Figure 20. Equation (58), on the other hand, has a 'stratified sampling' effect insofar as it adjusts the selection probabilities of the variables based on their performance. The benefit of this is partly that it 'accelerates convergence' (SDM, 2004: 819).

SDM use nine equations, including the one replicated above as equation (54), to explain their approach. Their equation (6), reproduced below, is a key one in terms of the mechanics described above.

$$\frac{P(M_0 \mid y)}{P(M_1 \mid y)} = \frac{P(M_0)}{P(M_1)} T^{(k_1 - k_0)/2} \left(\frac{SSE_0}{SSE_1} \right)^{-T/2} \qquad (60)$$

In contrast to the earlier equation (54), which referred to the prior and posterior distribution of parameter values, the above equation refers to the prior and posterior probability that the models M_0 and M_1 are the truest models explaining, in our case, the growth phenomenon. This extension of Bayesian statistics is a common one (Kennedy, 2008: 228). The left-hand side of equation (60) is the odds ratio, or the ratio of the first model's probability of being true to that of the second model. Here *k* refers to the actual number of explanatory variables, in other words what has been referred to as *m* in the previous discussion. *SSE* is *r* and *T* is *n* from equation (57). If equation (57) is used to calculate the odds ratio between two models, the result is identical to what one would obtain using equation (60).

How widely used is BACE? A few analysts have made use of the method. Jones and Schneider (2006), for instance, use results from intelligence or IQ tests as one of several explanatory variables and find that intelligence is associated with faster economic growth. This is less controversial than may at first seem to be the case if one considers that the intelligence data they use is essentially a measure of the quality of education. Nevertheless, the intelligence dataset used has been criticised by economists such as Ervik (2003) for, amongst other things, its lack of standardisation. Masanjala and Papageorgiou (2008) use BACE in an attempt to understand what lies behind the African dummy which SDM find important (it carries a posterior inclusion probability of 0.154 and is therefore amongst the top 18 variables discussed earlier). The analysis involves extensive use of interaction terms based on the SDM dataset and arrives at the conclusion that the negative impact of the African dummy is large due to an especially negative impact of a reliance on primary (as opposed to non-primary) exports in the case of Africa. However, they also find that in the case of Africa the level of primary enrolments and the fraction of GDP linked to mining had exceptionally positive effects in the case of Africa. Burger and Du Plessis (2006) also explore the dynamics behind the African dummy using BACE, but with some new data added to the SDM dataset, and find that certain omitted variables explain the African dummy, for instance changes to the dependency ratio.

Here the focus is on the impact of inserting an educational quality variable into SDM's original BACE analysis. To ensure that the Mata code used for the present analysis did not produce notably different results from the original GAUSS code, the original SDM analysis was re-run twice using the Mata code. Table 10 reports the results. The value 'differ' in the last row of the table is a measure of the degree to which the rankings of variables differ from those in the original SDM analysis. Zero would mean the exact same ranking was reproduced[44]. Any variable where the posterior inclusion probability was at least 0.104 (7 over 67) is listed. 'PIP' is posterior inclusion probability and 'Coeff' is the posterior mean conditional on inclusion, or the slope coefficient. The latter is given in scientific notation given the large range of coefficient values, which reflects the large range of standard deviations (see Table 12). The 'differ' values in Table 10 are low. Using the Mata code does not change the ranking or statistics to any notable extent. The differences between the two Mata results confirm what one should expect, given the random selection elements within the BACE process, namely that no two runs would produce identical results. Yet the results for at least the top seven variables can be considered highly stable.

Table 11 allows for an exploration of the impact of introducing a measure of educational quality. In this table, the first two panels report on the results of two analyses using only the 46 countries with data in the SDM dataset and the Hanushek and Woessman (2009) set of normalised country scores (see section 3.3 above). For the first panel, the country score ('cognitive') was included, in the second panel it was not. The second panel reveals that just the reduction in the number of countries, and fewer lower and middle income countries, alters the results in substantial ways (specifically, the percentage of lower and middle income countries drops from 72% to 48%). The number of explanatory variables with posterior inclusion probabilities not below the critical threshold of 0.104 increases, from 18 in the original SDM to 35 (see the second panel of Table 11). It has become less possible to distinguish between the importance of the variables. From an education policy perspective, a key finding is that the educational quality variable replaces the primary enrolments variable ('p60'), which no longer surpasses the critical threshold (see the first panel of Table 11). The BACE analysis thus arrives at similar conclusions to those of Hanushek and Woessman (2009), namely that to improve economic growth, policymakers should primarily focus on educational quality. High enrolments are a necessary but not sufficient condition for the realisation of tertiary educational quality. A similar outcome is seen in the third and fourth panels of Table 11, where for the analysis the new measure of educational quality described in section 3.3 ('newscore') was used (60% of the 60 countries in question are lower and middle income countries).

[44] Each value for 'differ', for instance 0.6 in the panel 'Replication of SDM 1', is the mean over a number of positive values. There is a positive value for every row where there is some content in the panel in question and the first panel 'Original SDM'. Each positive value is the absolute difference between the ranking of a variable in the first panel and in the panel in question. Thus for the first seven rows between the first two panels, the value is zero because the ranking of variables remains the same. For the eighth row, the value is 1 because 'life060' has ranking 8 in the first panel and ranking 9 in the second panel. For the last four rows, 'spain' to 'buddha', the value is 2 in each row.

One thing that is noteworthy is that whilst the primary enrolments variable takes position two in the Table 10 results, the educational quality variables in Table 11 take positions seven and five. An obvious explanation for this is that whilst the primary enrolments values are from 1960, in other words the start of the growth period considered in the dependent variable, the educational quality variables relate to what is essentially the end of the growth period: the average data source year in the case of Hanushek and Woessman (2009) is 1994 and for the new values data points from 2000 to 2010 were used. Had an educational quality variable for 1960 existed, this would perhaps have been allocated a ranking similar to the primary enrolments variable, given that learning outcomes in schools in 1960 would clearly be a better proxy for adult human capital in the 1960 to 1996 growth period, than a learning outcomes variable referring to the end of period.

What is also noteworthy is that some other variables considered highly important in Table 10 do not appear in Table 11. This can be said of the variables referring to the percentage of the country within the tropics ('tropicar') and the extent of malaria ('malfal66'). A possible explanation is the greater degree of multicollinearity between educational quality and these two variables, relative to that between primary enrolments and the two variables. Less easy to explain would be the presence of real exchange rate distortions ('rerd') in first or second position in Table 11, though it is absent in Table 10. The real exchange rate distortions variable, together with another two, do surpass a less stringent threshold in SDM (2004: 829), however. The fact that the variable is highly ranked in all the Table 11 panels is an indication that this is a question of the countries selected, not a question of the use of an educational quality variable.

The magnitude of the coefficients associated with the two educational quality variables in Table 11 is somewhat lower than what Hanushek and Woessman (2009: Table 1 in appendix) predict following their analysis (discussed below). This should be expected, given the larger number of variables considered in the BACE analysis. Specifically, an increase in the growth rate of 0.8 percentage points associated with an improvement of one standard deviation in the test scores is suggested by Table 11, against the comparable improvement of 2.0 percentage points suggested by Table 14 in the following section, which reproduces the original Hanushek and Woessman growth regression results (note that as the dependent variable used for Table 11 is expressed as a proportion, so a growth rate of 1 per cent is expressed as 0.01, coefficients from Table 11 must be multiplied by 100 to make them comparable to those in Table 14). Once-off regression analyses of the top seven variables in Table 11 result in coefficients which are consistent with the BACE values. The BACE approach thus assists in placing models such as those of Mankiw, Romer and Weil (1992) and Hanushek and Woessman (2009) within a wider development framework and in reducing the problem of over-estimation of the magnitudes of associations brought about by a narrow selection of explanatory variables.

Table 10: BACE regression outputs – without educational quality variables

Original SDM			Replication of SDM 1			Replication of SDM 2		
Var.	PIP	Coeff.	Var.	PIP	Coeff.	Var.	PIP	Coeff.
east	0.823	2.E-02	east	0.807	2.E-02	east	0.832	2.E-02
p60	0.796	3.E-02	p60	0.795	3.E-02	p60	0.798	3.E-02
iprice1	0.774	-8.E-05	iprice1	0.769	-8.E-05	iprice1	0.771	-8.E-05
gdpch60l	0.685	-9.E-03	gdpch60l	0.700	-9.E-03	gdpch60l	0.686	-9.E-03
tropicar	0.563	-2.E-02	tropicar	0.552	-1.E-02	tropicar	0.572	-1.E-02
dens65c	0.428	9.E-05	dens65c	0.424	9.E-06	dens65c	0.439	9.E-06
malfal66	0.252	-2.E-02	malfal66	0.246	-2.E-02	malfal66	0.255	-2.E-02
life060	0.209	8.E-04	confuc	0.222	5.E-02	life060	0.207	8.E-04
confuc	0.206	5.E-02	life060	0.211	8.E-04	confuc	0.194	5.E-02
safrica	0.154	-2.E-02	safrica	0.178	-1.E-02	safrica	0.155	-1.E-02
laam	0.149	-1.E-02	laam	0.172	-1.E-02	laam	0.150	-1.E-02
mining	0.124	4.E-02	mining	0.151	4.E-02	spain	0.124	-1.E-02
spain	0.123	-1.E-02	muslim00	0.127	1.E-02	mining	0.114	4.E-02
yrsopen	0.119	1.E-02	buddha	0.124	2.E-02	muslim00	0.110	1.E-02
muslim00	0.114	1.E-02	spain	0.120	-1.E-02	avelf	0.108	-1.E-02
buddha	0.108	2.E-02	yrsopen	0.108	1.E-02	yrsopen	0.107	1.E-02
avelf	0.105	-1.E-02				buddha	0.104	2.E-02
gvr61	0.104	-4.E-02						
n		88	n		88	n		88
iterations			iterations		54.6m	iterations		52.8m
differ.			differ.		0.6	differ.		0.5

Note: Dependent variable here and in the regressions of the following table is average real GDP per capita growth between 1960 to 1996.

Table 11: BACE regression outputs – effect of educational quality variables

H&W measure			Same observations			New measure			Same observations		
Var.	PIP	Coeff.	Var.	PIP	Coeff.	Var.	PIP	Coeff.	Var.	PIP	Coeff.
rerd	1.000	-2.E-04	rerd	1.000	-2.E-04	gdpch60l	0.968	-2.E-02	gdpch60l	0.963	-2.E-02
yrsopen	1.000	2.E-02	avelf	0.988	-4.E-02	rerd	0.944	-2.E-04	rerd	0.940	-1.E-04
avelf	0.999	-3.E-02	yrsopen	0.987	2.E-02	mining	0.788	5.E-02	life060	0.900	1.E-03
gdpch60l	0.990	-1.E-02	govnom1	0.983	-1.E-01	life060	0.782	1.E-03	confuc	0.801	5.E-02
govnom1	0.985	-1.E-01	gdpch60l	0.925	-1.E-02	newscore	0.755	8.E-03	mining	0.714	5.E-02
mining	0.952	4.E-02	mining	0.881	4.E-02	buddha	0.543	2.E-02	spain	0.686	-1.E-02
cognitive	0.866	8.E-03	spain	0.878	-2.E-02	confuc	0.515	4.E-02	safrica	0.665	-2.E-02
spain	0.782	-2.E-02	lhcpc	0.818	-7.E-04	spain	0.457	-1.E-02	buddha	0.639	2.E-02
pi6090	0.739	2.E-04	brit	0.732	8.E-03	dens65c	0.342	6.E-06	govnom1	0.326	-6.E-02
lhcpc	0.728	-6.E-04	newstate	0.602	4.E-03	avelf	0.330	-2.E-02	ggcfd3	0.309	-7.E-02
opendec1	0.571	6.E-03	europe	0.569	-1.E-02	safrica	0.318	-1.E-02	dens60	0.242	2.E-05
europe	0.520	-9.E-03	pop6560	0.432	2.E-01	tot1dec1	0.313	1.E-01	avelf	0.231	-2.E-02
colony	0.363	6.E-03	opendec1	0.417	9.E-03	govnom1	0.231	-7.E-02	priexp70	0.229	-2.E-02
geerec1	0.348	2.E-01	pi6090	0.415	2.E-04	east	0.188	1.E-02	dens65c	0.227	5.E-06
cath00	0.311	-6.E-03	cath00	0.403	-9.E-03	yrsopen	0.158	1.E-02	muslim00	0.208	1.E-02
brit	0.253	6.E-03	sqpi6090	0.402	2.E-06	priexp70	0.114	-1.E-02	opendec1	0.155	9.E-03
sqpi6090	0.253	3.E-06	confuc	0.284	2.E-02	ggcfd3	0.110	-6.E-02	ztropics	0.153	1.E-02
laam	0.247	-2.E-02	life060	0.266	5.E-04	dens60	0.108	2.E-05	europe	0.145	6.E-03
revcoup	0.245	-2.E-02	size60	0.206	3.E-03	dpop6090	0.108	-4.E-01	pop60	0.144	1.E-07
pop6560	0.234	1.E-01	geerec1	0.197	2.E-01				yrsopen	0.141	1.E-02
priexp70	0.211	-8.E-03	fertldc1	0.169	-1.E-02				east	0.134	1.E-02
confuc	0.190	2.E-02	scout	0.159	4.E-03				h60	0.134	-8.E-02
iprice1	0.162	5.E-05	prot00	0.156	-7.E-03				dpop6090	0.114	-5.E-01
newstate	0.161	3.E-03	dpop6090	0.150	8.E-01						
prot00	0.122	-2.E-03	govsh61	0.149	9.E-03						
dens65i	0.121	-1.E-05	ztropics	0.140	8.E-03						
pop1560	0.121	-4.E-02	east	0.140	7.E-03						
size60	0.110	1.E-03	oil	0.139	8.E-03						
			dens65i	0.130	-1.E-05						
			dens65c	0.127	-3.E-06						
			landlock	0.126	-7.E-03						
			revcoup	0.123	-1.E-02						
			laam	0.121	-1.E-02						
			tot1dec1	0.108	5.E-02						
			priexp70	0.106	-9.E-03						
n		46	n		46	n		60	n		60
iterations		52.4m	iterations		58.0m	iterations		84.4m	iterations		72.7m
differ.		20.8	differ.		21.7	differ.		9.1	differ.		11.3

Table 12: Variables reported on in BACE tables

		Non-missing	Mean	Std. dev.
East Asian dummy	east	139	0.115	0.320
Primary schooling in 1960	p60	121	0.692	0.321
Investment price	iprice1	119	94.354	53.074
GDP in 1960 (log)	gdpch60l	120	7.319	0.887
Fraction of tropical area	tropicar	119	0.570	0.467
Population density coastal in 1960s	dens65c	114	125.739	452.237
Malaria prevalence in 1960s	malfal66	115	0.379	0.438
Life expectancy in 1960	life060	127	53.129	12.277
Fraction Confucian	confuc	136	0.012	0.069
African dummy	safrica	139	0.317	0.467
Latin American dummy	laam	139	0.209	0.408
Fraction GDP in mining	mining	130	0.057	0.090
Spanish colony	spain	139	0.115	0.320
Years open 1950-94	yrsopen	126	0.334	0.347
Fraction Muslim	muslim00	136	0.203	0.346
Fraction Buddhist	buddha	139	0.038	0.155
Ethnolinguistic fractionalization	avelf	127	0.360	0.308
Gov. consumption share 1960s	gvr61	113	0.119	0.077
Population density 1960	dens60	126	155.285	415.449
Real exchange rate distortions	rerd	116	125.336	40.271
Fraction speaking foreign language	othfrac	133	0.297	0.395
Openess measure 1965-74	opendec1	125	0.570	0.354
Political rights	prights	134	4.099	2.020
Government share of GDP in 1960s	govsh61	119	0.165	0.072
Higher education 1960	h60	121	0.031	0.046
Fraction Population In Tropics	troppop	119	0.311	0.369
Primary exports 1970	priexp70	129	0.753	0.278
Public investment share	ggcfd3	111	0.055	0.045
Fraction Protestants	prot00	136	0.144	0.288
Fraction Hindus	hindu00	137	0.029	0.117
Fraction population less than 15	pop1560	119	0.396	0.072
Air distance to big cities	airdist	118	4401.263	2504.810
Nominal government GDP share 1960s	govnom1	119	0.148	0.060
Absolute latitude	abslatit	137	22.557	15.775
Fraction Catholic	cath00	136	0.298	0.399

Note: Variables are listed in descending order according to PIP in SDM. Statistics are based on the full 139 observation dataset of SDM.

4.4 Hanushek and Woessman's model

Hanushek and Woessman (2007, 2008, 2009)[45] have been particularly influential in policy circles, both in the education sector and in national planning generally. Not only have the authors presented findings that appear to provide a basis for greater certainty and specificity in policymaking, they also do so in a style that addresses policymakers directly, translating empirical findings to simple (though rudimentary) simulations of national trends that can guide actual planning.

The empirical analysis of Hanushek and Woessman (2009 – it is mostly this text that is considered in this section) is not strongly rooted in any single theoretical model. If one attempts to situate their work amongst the models from section 2, it becomes apparent that the role of physical capital *K* is not taken into account. Convergence, as we shall see, is predicted by Hanushek and Woessman's analysis. They attempt to analyse the separate effects of human capital of a basic and of an advanced quality. By introducing this focus, Hanushek and Woessman (2009) express concerns that are close to those of the Schumpeterian models, specifically that of Nelson and Phelps (section 2.7) and the Schumpeterian model with two levels of human capital (section 2.9). The absence in Hanushek and Woessman's work of a theoretical framework more explicitly rooted in the literature described in section 2 is cause for one of several criticisms made by Breton (2011), whose more salient criticisms will be discussed within this section. In a report produced for the OECD (2010a) by Hanushek and Woessman, one finds what is perhaps the clearest statement of the authors' theoretical underpinnings: they subscribe to endogenous growth theory, specifically the notion that a more educated country is better at innovating and therefore attaining a higher growth rate.

An important part of Hanushek and Woessman (2009) is the presentation of a novel way of combining country-level test score averages from different testing programmes. This was discussed in some detail in section 3.3. What is discussed below is how their new set of normalised 'country scores' was used within various growth models. The following table provides details on the data sources used for the replication of Hanushek and Woessman's (2009) work in this section. Maximum *N* is in all cases the 113 countries for which new learning outcomes values were calculated using the nonlinear programming approach explained in section 3.3.

[45] See also Hanushek and Wößmann (2007).

Table 13: Data sources used for replicating H&W growth regressions

Variable	*Source*	*N*	*Mean*	*Min.*	*Max.*
Real annual growth in GDP per capita (in PPP USD of 2006), average 1960-2004	Heston, Summers and Aten, 2006.	107	2.49	-2.00	8.36
GDP per capita (in real PPP USD of 2006), 1960	As above.	67	4753	412	15253
Gross enrolment ratio in secondary schooling, average 1980-2005	Online data-querying facility of UNESCO: UIS, consulted 2009.	101	67	4	118
Average country score	See discussion above.	113	427	260	578
Economy openness in 2000, following methodology of Barro and Sala-i-Martin (2003: 529).	Heston, Summers and Aten, 2006; Online data-querying facility of United Nations Statistics Division (National accounts estimates of main aggregate), consulted 2009.	110	5.7	-68.3	244.9
Rule of law indicator for 2000	World Bank, 2010.	107	0.30	-1.25	1.92
Total fertility rate, 1960	Online data-querying facility of United Nations Statistics Division (World population prospects: The 2008 revision), consulted 2009.	108	4.88	1.82	8.40

In the next table three basic models from Hanushek and Woessman (2009, Table 1, models (1) to (3)) are reproduced. The form of the most complete model (3) can be described as follows:

$$g_Y = B_0 + B_1T + B_2E + B_3Y_{t=0} + u \qquad (61)$$

Average annual growth in GDP per capita, g_Y, is a function of three variables: human capital T as seen in literacy and mathematics testing programmes, human capital E with respect to the degree to which the population has been enrolled in school, and initial income per capita Y at the start of the period. The form is completely linear. Moreover, growth is strongly predicted insofar as the growth rate, and not income itself, is the dependent variable. Thus once growth exists, it is not necessary for either of the two human capital variables to experience increases over time for further growth to occur.

The first three data columns of Table 14 display values copied from Hanushek and Woessman (2009). The basic conclusion that they draw from these values is that a growth model in which historical enrolments appear to be a statistically significant predictor of growth (this is the case in model (1)) is deceptive, as without a more qualitative measure of how educated the population is, the model suffers the specification error of an omitted yet relevant variable. Once the qualitative measure is introduced, historical enrolments become insignificant (see model (3)) and learning outcomes becomes a highly significant predictor. On its own, the learning outcomes variable is far more significant than historical enrolments on its own (compare models (1) and (2), including the very different adjusted R^2 values). The policy implications of this are profound. Policies in developing countries that focus exclusively on enrolling more pupils run the risk of devoting scarce national resources to a project which does not contribute to growth. Rather, governments should focus on what pupils learn. More years of schooling do not automatically translate into better learning outcomes as demonstrated, for instance, by the fairly low correlation between the two human capital variables *T* and *E* of 0.63 (using the data used for the right-hand panel of Table 14).

Table 14: Regression outputs – replication of H&W

	From H&W			*Replication with H&W values for T*		
Variable	*1*	*2*	*3*	*1A*	*2A*	*3A*
Constant				2.02***	-3.94***	-3.88***
				(3.52)	(-4.41)	(-4.30)
T [Learning outcomes]		2.02***	1.98***		1.82***	1.73***
		(10.68)	(9.12)		(8.67)	(7.26)
E [Enrolment]	0.37***		0.03	0.03***		0.01
	(3.23)		(0.34)	(3.23)		(0.81)
Y [Income]	-0.38***	-0.29***	-0.30***	-0.28***	-0.27***	-0.29***
	(-4.24)	(-9.15)	(-5.54)	(-4.54)	(-8.18)	(-6.85)
N	50	50	50	48	48	48
R^2				0.314	0.684	0.688
Adjusted R^2	0.252	0.733	0.728	0.284	0.670	0.667

*Note: Dependent variable is average annual growth 1960-2004. *** indicates that the estimate is significant at the 1% level of significance. Values in brackets are t statistics.*

Hanushek and Woessman use the term ‘cognitive skills’ to refer to abilities displayed through standardised tests. The term ‘learning outcomes’ used above is a term that might be more familiar to education planners. Here the two essentially mean the same thing. However, there is an important caveat in this regard that Hanushek and Woessman (2008: 612; 2009: 6) do acknowledge. Apart from the language and mathematical skills displayed in pupil tests, young people also acquire, partly through formal education, skills relating to social cohesion, for instance the skill to communicate well, or to suppress more selfish urges in the interests of the greater social good. Such skills are likely to be important for economic growth, yet they are not directly captured in test scores, though test scores and these other skills, called ‘noncognitive skills’ by Hanushek and Woessman, are likely to be correlated.

For the variable *E* Hanushek and Woessman (2009) use 1960 values from a dataset of average years of education produced by Cohen and Soto (2007b). These values are based on the total years of education of adults, including years of post-secondary education. They therefore deal not just with basic school education, which is the focus of the learning outcomes variable *T*. The problem is that overall years of education is in many ways a poor indicator of the level of participation in basic education. To illustrate, within the Cohen and Soto (2007b) dataset, Costa Rica and Malaysia have very different values for the percentage of adults with at least some primary schooling in 1960 (75% against 50%), yet the overall average years of education values for the two countries are almost identical (3.3 against 3.2 years). This is due to a stronger post-primary emphasis in the case of Malaysia. It would have been interesting to use a variable on the average years of pre-tertiary schooling of adults in the growth model, but such a variable is not available in Cohen and Soto (2007b) (nor does not it seem available anywhere else).

Clearly any enrolment variable will have its advantages and disadvantages. In models (1A) and (3A) of Table 14 the secondary level gross enrolment ratio (GER) in 1980 is used for variable *E*. This approach is thus similar to what was employed for the Mankiw, Romer and Weil (1991) simulations in section 4.2. The aim of the last three columns in Table 14 is to examine to what degree the Hanushek and Woessman (2009) conclusions are sensitive to the measure of historical enrolments used. It appears as if a different measure makes no substantial difference to the conclusions. Above all, *E* remains insignificant in model (3). The original Hanushek and Woessman (2009) learning outcomes values are used and for initial income in 1960 Penn version 6.2 data (Heston, Summers and Aten, 2006) are used (Hanushek and Woessman used Penn version 6.1 data, which is the reason why the number of countries drops from 50 to 48 in Table 14).

What is the effect of using the new learning outcomes values discussed in section 3.3? Simply rerunning the second model (2) in Table 14 using the new learning outcomes values results in a relatively low level of explanation reflected in an adjusted R^2 value of 0.519. The data allowed for 66 countries to be included in the model. If the 46 countries for which both new and old (or Table 14) learning outcomes values are available, then the adjusted R^2 value in model (2) of Table 14 is almost the same, regardless of which learning outcomes values are used (adjusted R^2 is 0.655 and 0.649 with the new and old values respectively).

Transforming the explanatory variables and excluding one outlier results in adjusted R^2 values that are comparable with those in the original Hanushek and Woessman (2009) models. The results appearing in Table 15 use the natural logarithm of both the enrolment values (secondary level GER) and the 1960 income values. A quadratic form for *T* was tested, but this brought virtually no benefits in terms of the value of adjusted R^2. The outlier that was excluded was Kenya. Potential outliers were identified using the DFITS post-estimation statistic (Baum, 2006: 128). In the case of Kenya, essentially the country's economic growth was much lower than what the model (and the level of their learning outcomes) would predict. China was excluded from all models as cognitive tests were only implemented in the city of Shanghai.

Table 15: Regression outputs – with transformed explanatory variables

Variable	*1*	*2*	*3*
Constant	4.21*** (3.60)	3.28*** (3.56)	3.87*** (4.65)
T		1.88*** (9.18)	1.54*** (7.74)
ln(*E*)	1.53*** (5.66)		0.87*** (4.11)
ln(*Y*)	-0.96*** (-4.59)	-1.08*** (-6.90)	-1.40*** (-8.78)
N	63	63	63
R^2	0.350	0.585	0.678
Adjusted R^2	0.329	0.572	0.661

*Note: Dependent variable is average annual growth 1960-2004. *** indicates that the estimate is significant at the 1% level of significance. Values in brackets are t statistics. For this analysis and those that follow the new learning outcomes values are divided by 100 to provide slope coefficients more or less comparable to those of Hanushek and Woessman (2009).*

The Table 15 results permit similar conclusions about education and growth to those drawn by Hanushek and Woessman (2009), though there is a noteworthy difference. Models (1) and (2) confirm how much more valuable educational quality, as opposed to education enrolment, is if we want to explain why certain countries experience higher economic growth than others. In model (3) both educational quality and enrolment are highly statistically significant, whilst in the comparable Hanushek and Woessman (2009: Table 1, Model 3) model only educational quality was found to be significant. This is mainly due to the fact that more developing countries have been included. If model (3) above is run excluding all countries excluded by the original Hanushek and Woessman analysis, the enrolment variable becomes statistically insignificant. The fact that the enrolment variable has been logged in model (3) is of little importance. If the unlogged form is used, the variable remains statistically significant (above the 10% level of significance). The coefficients associated with initial income remain negative and statistically significant across all models, something that Hanushek and Woessman (2009: 9) remind us points towards conditional convergence between countries.

In a model with a wider range of explanatory variables, roughly following Hanushek and Woessman (2009: Table 1, Model 8), educational quality as reflected by the new country scores remains statistically significant. The adjusted R^2 value in Table 16 of 0.79, in a model that includes 62 countries, compares to an adjusted R^2 value of 0.80 in the corresponding Hanushek and Woessman model with 45 countries. The most significant regional dummies were those of Sub-Saharan Africa (SSA) and Middle East and North Africa (MENA), both of which produced strongly negative coefficients. Below, the two are combined into one dummy variable. Following Burger and Du Plessis (2006), it is likely that this dummy variable masks omitted variables relating to, for instance, changes in the dependency ratio. With respect to the dependency ratio, these authors found that poor economic performance in African countries can be explained by an absence of the demographic trends, such as a lower fertility rate, needed to increase the labour force as a proportion of the total population. What is also confirmed by Table 16 is the finding, evident in Hanushek and Woessman (2009) though not emphasised by them, that a more comprehensive model reduces the magnitude of the impact of educational quality on growth by about half: a coefficient of 1.54 in Table 15 compares to 0.70 in Table 16.

Table 16: Regression outputs – large range of explanatory variables

Variable	
Constant	9.92*** (6.53)
T [Learning outcomes]	0.70*** (3.02)
ln(*E*) [Enrolment]	0.85*** (4.65)
Openness	0.01*** (3.31)
Rule of law	0.43** (2.34)
Total fertility rate	-0.13 (-1.49)
SSA/MENA dummy	-0.53** (-2.15)
ln(*I*) [Income]	-1.62*** (-10.36)
N	62
R^2	0.794
Adjusted R^2	0.768

*Note: Dependent variable is average annual growth 1960-2004. *** indicates that the estimate is significant at the 1% level of significance, ** at the 5% level. Values in brackets are t statistics.*

To conclude the preceding discussion, there appear to be two findings of note arising out of the use of a somewhat enlarged set of countries with respect to learning outcomes, where this enlargement is made possible through the incorporation of a couple of additional international testing programmes. Firstly, the essential findings around the importance of learning outcomes by Hanushek and Woessman (2009) still stand. Secondly, the importance of enrolments when viewed beside learning outcomes cannot be discarded completely, especially if a larger number of developing countries are included in the model.

Hanushek and Woessman (2009: 22) attempt to examine the separate growth effects of two different variables: the proportion of the population achieving a basic level of learning outcomes, and the proportion achieving exceptional results. They create two variables on the percentage of pupils achieving above a specific standard. 86% of pupils internationally achieve at or above their lower standard, whilst 5% achieve at or above their higher standard. The two variables are found to be simultaneously significant when run in the same growth regression. The analysts conclude that policymakers should consider both widespread basic achievement and a critical level of exceptional performance as policy aims at the same time. It is not a matter of choosing between the two, in particular considering the two appear to reinforce each other.

A central concern in Hanushek and Woessman (2009: 13) is the possibility that the stylised facts seen in the kinds of cross-country regressions discussed above may be deceptive because of endogeneity biases within the models. This possibility, they admit, was compellingly argued by in particular Bils and Klenow (2000), whose analysis suggested that less than a third of the correlation between initial levels of educational enrolment and subsequent economic growth can be attributed to a causal impact of education on growth. Instead, Bils and Klenow (2000) find evidence that anticipation of growth prompts households to invest in more schooling in order to reap future benefits. A strong reverse causality is thus suggested. Hanushek and Woessman (2009) argue that through their own analysis they sufficiently lay to rest any doubts about the strength of the education effect on growth through a verification process involving the use of instrumental variables. Metaphorically, they stress test their regression results using instrumental variables (IVs).

Formally, endogeneity is said to exist when an independent variable, such as H in equation (61), is correlated with the error term, such as u in equation (61). This means that what Gujarati (2003: 71) identifies as assumption 6 of his 10 assumptions underlying the unbiased classical linear regression model is violated. According to this assumption, there should be zero covariance between the error term and each independent variable. Endogeneity points to a bias in the slope coefficient of the independent variable in question, such as B_1 in equation (61). Endogeneity often arises due to one or a combination of the following three problems: an omitted variable, measurement error, and simultaneity (Kennedy, 2008: 138). The first two problems are of course in the list of specification errors introduced in section 4.1. Hanushek and Woessman (2009:13), in their discussion of endogeneity, are largely concerned with two of the three problems, namely omitted variable bias and simultaneity.

With respect to the omitted variable problem, it is possible that some variable not included in Table 16 might change the existing coefficients if added. An obvious example of an omitted variable in this instance is one referring to a cultural trait within the country. Hanushek and Woessman (2009: 13) suggest that 'cultural factors' might cause certain countries to produce both better school results and better economic growth. If this were true, it could be deceptive to regard the models discussed above as a basis for advising governments to gear education policies more towards achieving learning outcomes. Governments of countries with the 'wrong' cultural traits may not see growth improvements flowing from education policy changes. Instead, the appropriate advice might rather be to change the national culture somehow.

Simultaneity means there is codetermination between the dependent variable and an explanatory variable. For instance, better learning outcomes may be contributing to economic growth whilst simultaneously better economic growth results in greater spending on education and hence better learning outcomes. The following two equations illustrate the existence of simultaneity.

$$Y_i = B_0 + B_1 X_i + B_2 Z_i + u_i \tag{62}$$

$$Z_i = A_0 + A_1 X_i + A_2 Y_i + v_i \tag{63}$$

The above situation, where Y (for instance economic growth) is a function of Z (for instance learning outcomes) whilst Z is also a function of Y necessarily results in Z being correlated to the error term u, or endogeneity. This can be demonstrated mathematically.

An instrumental variable is formally a new variable that is correlated with the existing independent variable suspected of being endogenous – for example T in equation (61) or Z in equation (62), but also not correlated with the error term (Gujarati, 2003: 527). More intuitively, an instrumental variable captures the part of the endogenous variable that represents the effect one is interested in from a theoretical or economic perspective. For instance, if one is interested in the impact of smoking on lung cancer, then having cigarette prices as an instrumental variable can help one to isolate the impact one is interested in, and ignore other dynamics, such as the possibility that people get lung cancer from living in polluted cities, whose cultural characteristics also make people smoke more. If the incidence of lung cancer diminishes when the price of cigarettes rises, one can probably be very certain that cigarettes cause lung cancer. Hanushek and Woessman (2009) select several instrumental variables reflecting the structural properties of the education system, such as the percentage of pupils in the schooling system who take standardised examinations. Such variables are likely to influence test results, but not economic growth in any direct way. By using these instrumental variables, they are able to isolate the part of the relationship between test scores and economic growth that is attributable to an education effect, in much the same way as using cigarette prices allows one to isolate the causal link between smoking and lung cancer.

The instrumental variable approach, implemented using the ivregress command in Stata, involves two stages of analysis. In the first stage, all non-endogenous (or exogenous) explanatory variables plus the instrumental variable or variables are regressed against the variable suspected of being endogenous. This analysis involves a straightforward OLS regression. In the second stage, all exogenous and endogenous variables, but not the instrumental variables, are regressed against the original dependent variable (for instance average annual GDP growth). These two stages are discussed below.

The Stata command ivregress has three estimation techniques, the most commonly referred to one being two-stage least squares, or 2SLS, indicated by ivregress 2sls in Stata. It is this technique which was used for the analysis discussed below.

An instrumental variables analysis can be explained by comparing it to a regular OLS regression in terms of matrix algebra. An OLS regression is represented as follows (Gujarati, 2003: 933):

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{64}$$

In contrast, the second stage of the 2SLS analysis is represented by the next two equations (Baum, 2006: 188):

$$\hat{\boldsymbol{\beta}} = (\mathbf{M}'\mathbf{X})^{-1}\mathbf{M}'\mathbf{y} \tag{65}$$

In two of three places the matrix of explanatory variables X is replaced by the matrix M, which is obtained as follows:

$$\mathbf{M} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X} \tag{66}$$

We can think of X as having the dimensions $n\times(k_1+k_2)$, where k_1 is the number of explanatory variables that are regarded as exogenous and not suspected of being endogenous whilst there are k_2 variables which we suspect could be endogenous. The matrix Z has the dimension $n\times(k_1+z)$, where z is the number of instrumental variables. The number of instrumental variables, or z, must be greater than or equal to the number of endogenous variables, or k_2. Importantly, equation (65) and equation (66) do not represent the two stages of the 2SLS analysis. Both equations together provide the second-stage results and clearly do not represent a typical one-stage OLS regression analysis. The first-stage estimated coefficients are provided by the following equation (which does represent an OLS regression):

$$\hat{\boldsymbol{\alpha}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{w} \tag{67}$$

Here w is the $n\times1$ column vector representing the one endogenous variable being tested. Only one such variable can be tested at a time. The outputs of the first stage of the analysis are useful as background information but are not essential to understand what the instrumental variables analysis finds with respect to the endogenous variable in question. In fact, the display of the first-stage OLS results is only an option in the ivregress command.

Hanushek and Woessman (2009), in each of four different instrumental variables analyses, use one or two instrumental variables out of a set of four instrumental variables they identify as possibly useful. Their four instrumental variables are: percentage of pupils subject to standardised examinations, extent of private schooling, degree of autonomy enjoyed by schools, and extent of school enrolment. The last variable is the same as one of their original explanatory variables. Hanushek and Woessman (2009) argue that not only are their instrumental variables likely to be correlated with their country test scores variable, they are moreover likely to be embedded in historical and legal traditions and thus *not* likely to be correlated to growth or the error term. In the second-stage outputs of the instrumental variables analysis, the slope coefficient for the learning outcomes variable remains more or less what it was in the original OLS regression and the coefficient remains significant at the 1% level (though the *t* values are lower than what they were in the corresponding OLS regressions). On the basis of these results, combined with the results of a Durbin-Wu-Hausman test examining the validity of the IV results, they conclude that policy-induced improvements in learning outcomes are indeed causing better growth. Whilst improvements in learning outcomes originating in the growth itself or cultural factors may exist, those dynamics do not explain the whole situation. The data support the hypothesis that governments can improve economic growth by ensuring that the right education policies are implemented.

One can use an IV analysis to test the hypothesis that cultural factors play a role in promoting growth through better education. This one can do by using a variable describing cultural factors, as opposed to variables describing factors induced by educational policies, as the instrumental variable. One apparently relevant cultural variable is the indicator of cultural secularity calculated by Inglehart and Welzel (2005) using World Values Survey data. One might expect a more secular and less traditional culture to stimulate more intellectual enquiry and hence cognitive skills in the schooling system. Table 17 provides the results of an instrumental variables regression where the learning outcomes variable is instrumented by the secularity variable. The results of the first-stage regression indicate that there is indeed a statistically significant and positive correlation between learning outcomes (the dependent variable in the regression) and secularity, even when enrolment and initial income are entered as conditioning variables. The second-stage instrumental variables regression demonstrates that even when only learning outcomes differences explained by secularity and the other variables in the first-stage regression are taken into account, learning outcomes displays a significant association with growth. The analysis presented here thus indicates that even if educational policy factors influence growth through better education, at the same time some of the growth-inducing effects of educational performance stem not from educational policy directions, but deeply-rooted values held in the society. To compare, if the second-stage part of the IV regression of Table 17 had been run as a single-stage OLS regression, learning outcomes would have carried a *t*-value of 7.75 (see Table 18), instead of 4.76 (see Table 17). This points to the ‘stress test’ aspect of the IV regression. Essentially it is ‘more difficult’ for the variable in question to remain a significant predictor within the IV regression. However, it is not always true that the *t* statistic of the variable in question will be higher in the OLS regression than in the IV regression. Some experimentation with a variety of plausible variables will quickly reveal that this pattern does not always hold.

Table 17: Regression outputs – culture as an instrumental variable[46]

Variable	
First-stage regression (dependent variable: Learning outcomes)	
Constant	2.32*** (3.40)
ln(*E*) [Enrolment]	0.25* (1.79)
ln(*I*) [Income]	0.14 (1.45)
Secularity	0.31*** (4.35)
N	53
R^2	0.633
Adjusted R^2	0.611
Instrumental variables (2SLS) regression (dependent variable: Average annual growth)	
Constant	3.30*** (3.63)
Learning outcomes	1.78*** (4.76)
ln(*E*) [Enrolment]	0.87*** (3.17)
ln(*I*) [Income]	-1.46*** (-8.27)
R^2	0.708

*Note: Other than the secularity variable, variables are those used in the models displayed in Table 14 (learning outcomes is entered in the linear form here however). *** indicates that the estimate is significant at the 1% level of significance * at the 10% level. Values in brackets are t statistics.*

Table 18: Regression outputs – 2nd stage with simple OLS

Variable	
Constant	3.41*** (3.71)
Learning outcomes	1.58*** (7.75)
ln(*E*) [Enrolment]	0.97*** (4.05)
ln(*I*) [Income]	-1.41*** (-8.58)
N	53
R^2	0.714
Adjusted R^2	0.697

*Note: Dependent variable is average annual growth 1960-2004. *** indicates that the estimate is significant at the 1% level of significance. Values in brackets are t statistics.*

It should be added that if the secularity variable were to be added as an explanatory variable within the Table 18 regression, the learning outcomes variable remains significant at the 1% level, whilst the secularity variable is not significant even at the 10% level (t equals 0.62). Again, the robustness of learning outcomes as a predictor of growth seems to be confirmed.

There are, however, noteworthy caveats regarding the above instrumental variable analysis that emerge if one applies a few of the available post-estimation tests applicable to this type of analysis. As mentioned above, Hanushek and Woessman (2009: 15) apply the Durbin-Hausman-Wu test and use the results to conclude that their variable for educational quality can be considered exogenous and not endogenous. In other words, educational quality is indeed a causal factor behind faster

[46] Strictly speaking, the 't statistics' (the statistics in brackets) in the instrumental variables regression panel in Table 17 are z-values, but for the purposes of this discussion they can be considered t-statistics. In fact, Stata has been somewhat ambiguous here. Whilst in Stata 9 the significance statistics in the IV regression results were referred to as t-statistics, in Stata 12 statistics of a similar magnitude were referred to as z-values. Stata 12 was used for the Table 17 results.

economic growth. The manuals for Stata 12 recommend the use of the Durbin and Wu-Hausman tests (described as two separate tests) to establish whether the variable being considered possibly endogenous in the instrumental variable analysis is in fact exogenous. The manuals moreover suggest using a criterion put forward by Stock, Wright and Yogo to establish whether an instrumental variable is sufficiently correlated with the possibly endogenous variable (StatCorp, 2011: 830-3). The secularity instrumental variable does in fact appear sufficiently correlated to educational quality according to the Stock-Wright-Yogo criterion. Specifically, the F statistic and eigenvalue for secularity in the first stage regression shown in Table 17 found after the `estat firststage` post-estimation test exceeds 10 (the value in the case of both statistics is 18.9). Moreover, the p statistics in both the Durbin and Wu-Hausman tests, of 0.52 and 0.53 respectively, indicate that the null hypothesis that the educational quality variable is exogenous cannot be rejected. This is broadly in line with the finding of Hanushek and Woessman (2009).

Returning to what by 2011 was probably the most cogent critique of Hanushek and Woessman's work, a key argument made by Breton (2011) is that Hanushek and Woessman make inappropriate use of test score data in their growth regressions . Breton (2011: 768) implicitly criticises Hanushek and Woessman for not being more transparent about the years from which the test score data are obtained[47]. For this dissertation, the mean year of all the test score data points was calculated, as discussed in section 3.3, and was found to be 1994. There was thus a concentration of data points in the latter part of the 1964 to 2003 period considered by Hanushek and Woessman. This important point was not discussed in Hanushek and Woessman (2009), though most of the information needed was available in a table. The point is important because, as Breton (2011) argues, test score data from the schooling system are likely to reflect the quality of human capital in the labour market as it will be around 20 years after the data were collected. This implies that the data used for T in equation (61) in the preceding analysis are not able to predict the growth rate for the period of 1960 to 2000 (as is implied by Hanushek and Woessman (2009)). Breton also criticises Hanushek and Woessman's measurement of educational quantity and the functional form of the growth regressions, but his argument relating to the test score data is particularly compelling. It is even more applicable to the regressions reflected in Table 15 as these make use of even more recent test score data than those of Hanushek and Woessman.

To what extent does Breton's (2011) argument explained here invalidate the Hanushek and Woessman approach to understanding the relationship between education and economic growth? Importantly, in one variation of model (3) from Table 14 run by Hanushek and Woessman (2009, Table 2, model 9) for just 22 high income countries, almost certainly nearly all high income countries, where country test score averages are all from 1984 or an earlier year and the growth period is 1980 to 2000, the slope coefficient for the learning outcomes variable was found to be almost twice what was seen in the left-hand panel of Table 14. This is important and to a large extent deals with Breton's (2011) complaint. It is particularly important as one might expect a model with only high income countries to yield a lower slope

[47] The foregoing discussion on Hanushek and Woessman's work draws mainly from Hanushek and Woessman (2009). Breton's critique is directed at Hanushek and Woessman (2008). However, the two Hanushek and Woessman sources mentioned here are very similar so Breton's critique can be considered a critique of the modelling discussed previously in this section.

coefficient for educational quality, if one assumes diminishing marginal returns for this input in the growth process. The time periods for the 22-country model mentioned here almost satisfy Breton's (2011) requirement in the sense that 2000, the end of the growth period, is 16 years after 1984, the last test. However, Hanushek and Woessman (2008 and 2009) do not seem to address Breton's (2011) valid concern around historical periods and causality explicitly and are not sufficiently clear about the importance of the 22-country model, so Breton's (2011) critique remains valid. The clearest acknowledgement by Hanushek and Woessman that could be found (in 2012) of the historical periods and causality problem was in their report for the OECD (2010a: 15), but even here the extent of the problem and what could be done to overcome it could have been explained in more depth. Even if the 22-country model is a theoretically more defensible model, the question still remains whether one would obtain similar results if one also had the required data for developing countries and thus could run a model that was more representative of the world. One key thing that Hanushek and Woessman do point out in OECD (2010a: 15) and that Breton (2011) could have been expected to acknowledge, is that the data problems discussed here are likely to result in under-estimates of the impact of educational quality on growth so estimates can in some ways be considered minimum values. Moreover, the extent to which test score averages change over the course of some decades within a country ought be dealt with explicitly. The less change there is, the more defensible it is to use lagged measures of educational quality. One can conclude that despite these caveats, Hanushek and Woessman (2009) remains a milestone analysis on the relationship between education and growth insofar as it represented by far the most ambitious and clearly explained attempt to date to include educational quality within a growth regression, and the coefficients on educational quality were found to be large and significant.

4.5 Conclusion

Section 4 has dealt with econometric issues and policy issues. Apart from the latter, education policymakers must deal with the former to some extent, partly because they sometimes commission econometric work and partly because some familiarity with the available techniques is required for the conversation that must happen between economists and policymakers.

The work by Sala-i-Martin, Doppelhofer and Miller (2004) discussed in section 4.3 underlines an important point regarding data, namely that the available data include a wide variety of country-level variables that can used for understanding the social, cultural, institutional and geographical aspects of development. This point may be missed, given that the methods used by most econometricians preclude a consideration of a wide variety of variables. Overlooking the richness of the data that have become available limits both the discussion around the modelling alternatives and the depth of the policy conversation.

Theory should probably inform econometrics to a greater degree. Mankiw, Romer and Weil (1992) seem to offer a good example of how to proceed in this regard. Not only should theory inform the structure of the empirical equations, it should also inform what countries one excludes from one's growth regressions. Acknowledging, for instance, that economies based largely on oil are likely to display atypical growth patterns is important for the econometrician, but also for some more general debates, for instance those around whether Venezuela offers a model for countries such as South Africa when it comes to poverty reduction.

The Hanushek and Woessman (2009) finding that a test-based educational quality variable so easily eliminates the significance of years of educational attainment, and thus raises doubts around earlier growth modelling, has understandably been embraced by policymakers and institutions such as the World Bank (Hanushek and Wößmann, 2007). The finding, not really possible previously due to data limitations, has lent additional power to the argument, which many educationists have made for decades, that simply going through the motions of providing a school building, teachers, books and so, is often not enough. Such measures on their own are no guarantee that education and national development will happen. They need to be accompanied by monitoring and remediation that deal with the substantial risk that there might be little actual learning happening in schools. As has been demonstrated above, the Hanushek and Woessman (2009) finding is easily replicated when the educational quality variable is inserted into the models of other analysts.

The excitement arising from Hanushek and Woessman's (2009) work, however, has detracted from the serious specification errors of their empirical model discussed above. Above all, there is an obvious problem inherent in the use of data on the quality of secondary schooling which are largely from the 1990s as a proxy for the human capital of adults in the 1960 to 2000 period. That this anomaly has been so easily overlooked seems indicative of how necessary a sufficient degree of research vigilance is. The inconvenient reality is that we will probably need to wait for many more years of testing to occur before more rigorous cross-country analysis of the linkages between educational quality and growth is possible. In the meantime, findings such as those by Hanushek and Woessman (2009) need to be considered tentative, no matter how intuitively appealing they are, whilst they remain important, arguably far more important than growth models which have used participation in schooling as a proxy for the learning outcomes of schooling.

One important matter that Hanushek and Woessman (2009) do acknowledge and that policymakers need to pay attention to is that the size of the association between education and growth depends on how many explanatory variables are entered in the model. This was confirmed when the BACE findings of Sala-i-Martin, Doppelhofer and Miller (2004) were discussed. It seems advisable for policymakers to consider as more realistic the lower magnitudes of association found in the more inclusive models. Put differently, education policymakers need to acknowledge that development is a complex process that is also dependent on what their policymaker colleagues do, or do not do, in other areas such as health and justice.

The intuition of most education policymakers is unlikely to sit well with the apparently well argued texts of sceptics who have insisted, as in the case of Pritchet (2001), that there is no association between education and economic growth or, as in the case of Bils and Klenow (2000), that though there is an association, causality runs largely from growth to education, and not vice versa. Yet education policymakers are likely to run into these arguments from time to time and can benefit from some of the research, much of it compelling, that the intuition of the policymaker that improving education contributes to economic development is right. Key here is Krueger and Lindahl's (2001) point that insufficient attention to measurement error lies behind many of the findings of the sceptics. Hanushek and Woessman (2009), despite the shortcomings mentioned above, offer sound advice on how the direction of causality between education and growth can be tested. Despite the historical misalignments in their data, their finding that causality runs mostly from education to growth seems believable.

5 AUGMENTATIONS TO INFORM EDUCATION POLICYMAKERS

5.1 Introduction

As explained in section 5.2 below, the conversation that should happen between researchers, such as academic economists, and policymakers is much weaker than it ought to be. As discussed below, there is reason to believe that this problem is especially pronounced in the area of education policy. Part of the problem is that not enough attention goes towards synthesising research findings, in the form of meta-analyses, and presenting data and findings in ways that would 'talk to' policymakers, in other words in a way that acknowledges that policymakers tend to draw from many disciplines and can find the language and representational intricacies of single disciplines, such as economics, difficult to follow. Section 5 takes the data and models discussed in previous sections and considers how they can be presented and augmented to maximise their utility for policymakers. The education policy issues that receive attention are inevitably just a small sub-set of the larger set of all education policy issues. Yet, whilst the policy issues may be limited, the points made below about how economic research can be presented to policymakers are fairly broadly applicable.

The generally poor state of the conversation between researchers and policymakers, and more specifically economists and education policymakers, is examined in section 5.2. Section 5.3 turns to an important and under-explored area, namely the speed at which educational quality improvements can realistically be expected to occur. In a context where policymakers are increasingly expected to set targets, this is a crucial area of empirical analysis. In section 5.4 the focus is on how country-level development indicators, such as the variables used by the three empirical models discussed in section 4, can be examined to provide policymakers with a picture of how large the education challenge is, relative to other development challenges. Section 5.5 draws from the empirical findings discussed in section 4 and looks at ways of representing the economic growth impact of educational improvement relative to the impact of other policy interventions. Finally, section 5.6 explores ways in which policymakers can use available empirical findings to reinforce long-term understandings between the state and teacher unions in relation to matters such as teacher pay and educational improvement.

Sections 5.3 to 5.5 draw from an Excel-based tool, called the growth simulation tool here, which was developed to accompany this dissertation[48]. It is important to realise that the main aim of the tool is not to analyse, in the sense of finding new and relevant trends, but to synthesise and represent findings already established in the literature in a form that is accessible to a wider audience that includes policymakers. Excel is particularly suited for the latter aim.

Where applicable, the following sub-sections acknowledge work that has already occurred to make the education-growth linkages clearer to policymakers. Two fairly similar texts, both produced by Hanushek and Woessman, one for the World Bank and another for the OECD[49], stand out (Hanushek and Wößmann, 2007; OECD,

[48] File name is *Growth simulator.xls.*
[49] Organisation for Economic Co-operation and Development.

2010a). To some extent, the focus is on demonstrating how such work can be expanded upon.

5.2 A problematic research-policy nexus

The impact of research on policymaking is widely believed to be sub-optimal. Much of the considerable literature that has emerged to deal specifically with this topic takes the position that this impact is in some respects close to negligible and that this represents a huge cost to society, largely because the result is sub-optimal policies. A prominent early researcher dealing with the topic of the 'research-policy nexus', a term used by Bailey (2010), is Weiss (1979), who describes a number of models, some idealisations and others rather cynical, of how research relates to policymaking. For example, the 'knowledge-driven model' describes what is widely considered the ideal. Researchers focus to a large degree on matters that have indirect or direct relevance for human development. Policymakers understand the work of the researchers and take it into account when designing policies. In contrast, within the 'political model', ideology shapes policymaking to a large degree and research is used to legitimise policy decisions, if research can be found that is compatible with pre-determined policy positions. Amongst the more compelling explanations for the poor impact of research on policy is the 'two communities theory' put forward by, for instance, Ginsburg and Gorostiaga (2001). According to this theory, researchers and policymakers are so absorbed by their concerns within their own institutions, and are institutionally so divorced from each other, that they find little time to learn about each other's capabilities, needs and work. Researchers of the research-policy nexus problem have tended to contradict common wisdom and have called for changes in the way researchers work, in the manner of Weiss (1979: 431):

> There has been much glib rhetoric about the vast benefits that social science can offer if only policy makers paid attention. Perhaps it is time for social scientists to pay attention to the imperatives of policy-making systems and to consider soberly what they can do, not necessarily to increase the use of research, but to improve the contribution that research makes to the wisdom of social policy.

There is some literature on the relationship between economic research with an education focus, on the one hand, and education policymaking, on the other, and it tends to portray a relatively dysfunctional relationship. Psacharopoulos (1996: 353), an economist, has lamented that '[i]n the field of education, perhaps more than in any other sector of the economy, politics are substituted for analysis'. In a sense, the political model prevails. On the education side, there has been suspicion amongst education researchers, some of which would be shared by education policymakers, towards the true motives of economists dealing with education issues. The following extract from Resnik (2006: 176), appearing in the *Comparative Education Review,* expresses an extreme version of this suspicion:

> … the success of the education-economic growth approach has to be understood not in terms of its inherent validity but, rather, as a consequence of the ability of the actors identified with it to build a large and influential network. Both econometric economists of education and functionaries of UNESCO and the OECD participated in the construction of the black box.

The 'education-economic growth black box' is, according to Resnik, a fallacious theory claiming that an important function of education is to contribute towards economic growth. She is thus referring to the theory that this dissertation is largely concerned about. Reservations amongst educationists towards the human capital model were discussed in 2.1.

McCloskey (1994: 30), along the lines of Weiss, suggests that economists should pay closer attention to communicating in ways that make sense to policymakers:

> ...why doesn't social policy work? Sure, sure: the politicians and bureaucrats don't carry it out. But there's something wrong, too, with our conversation as economists if we can't advise them, and teach each other, in more persuasive ways.

As part of an effort to communicate the importance of communication, McCloskey (2008) offers a rough estimate of the proportion of the GDP of the United States that is attributable to persuasion. This proportion is said to be as high as 25%.

If one wants evidence that persuasion around the importance of improving educational quality was not a completed task by around 2010, the UNDP's 2010 Human Development Report, focussing specifically on human capital, is an interesting case. Surprisingly, in discussing education policy this report limits itself largely to issues such as education attainment, with scant acknowledgement of important work that had occurred during the previous decade or so in relation to standardised assessment systems and the contribution of educational quality to growth.

The analysis and discussion presented in the remainder of section 5 are aimed at taking the empirical and theoretical work covered in previous sections and constructing persuasive arguments that might influence education policymakers, or make it easier for them to influence others, for instance people in the Ministry of Finance or teacher union members.

5.3 Feasible education improvement targets

The spread of standardised testing across schooling systems has understandably led to a tendency amongst policymakers to set future targets for the level of learning outcomes. As will be seen below, targets that have been set are often not realistic. The empirical basis on which targets can be set is not clear in the literature. Exploring a better empirical basis is the principal concern of what follows. The discussion here is partly reflected in Gustafsson (2012a).

In the United States, as part of the No Child Left Behind policy, every state was expected to achieve certain targets. Hawaii, for example, produced a set of targets for the percentage of pupils attaining proficiency in mathematics which essentially envisaged a movement from a baseline of 10% in 2001 to 100% in 2014 (Hawaii: Department of Education, 2003). In other words, Hawaii's targets focussed on a percentage of pupils, not an overall average score. Brazil serves as an example of a country that has set targets based on average scores. In Brazil such targets at the federal, state, municipality and school level have been set using a central database of baseline scores and a formula that assigns higher targets where the baseline scores are higher[50]. Brazil's targets are based more or less but not exactly on average scores. The targets are in fact index values that include the average score and a grade promotion ratio. Combining these two variables within the index is intended to minimise the risk that schools will game the system by forcing weaker pupils to repeat their grades and

[50] See http://sistemasideb.inep.gov.br/resultado/.

hence delay (or even prevent, through dropping out) their participation in the test (Fernandes, 2007). South Africa aims to increase the percentage of primary school pupils performing at a basically acceptable level in languages and mathematics from a range of between 20% and 50% (depending on grade and subject) in 2010 to 60% in 2014, with monitoring occurring through the country's new Annual National Assessments programme (South Africa: Department of Basic Education, 2010a).

Some guidance regarding the possible speed of improvement in standardised test scores exists, but the basis for this important guidance is not sufficiently clear and some of the available guidance is contradictory. Hanushek and Woessman (2007: 44) indicate that in a programme such as TIMSS, an improvement of half a standard deviation (meaning 50 points in TIMSS) is possible within a period of 10 to 30 years. It is indeed practical to express possible improvements in terms of standard deviations in order to facilitate comparison across different testing systems testing using different scoring procedures. However, it is important to differentiate between a standard deviation across the test-takers of many countries in a programme and a standard deviation within one country. With regard to the former, both TIMSS and PIRLS establish by design a standard deviation equal to 100 points. Within-country standard deviations tend to be lower than this, for instance the median within-country standard deviation for grade 8 mathematics in TIMSS 2007 is 85 (Mullis, Martin and Foy, 2008: 457). One would expect within-country standard deviations to be smaller than the standard deviation derived from pupil observations across a large range of countries. However, a few countries do have an internal standard deviation that is greater than 100 in TIMSS. The Hanushek and Woessman guidance referred to above translates into between 0.02 and 0.05 standard deviations a year, assuming a linear trend over the years in question. If one were expressing the improvement in terms of a national standard deviation, something one may wish to do if one is comparing, say, TIMSS results to the results from a national testing system, the 0.02 to 0.05 parameters would probably need to be raised slightly, depending on the country. The parameters put forward by Hanushek and Woessman (2007) are slightly more conservative than what the analysis that follows suggests should be possible.

Mourshed, Chijioke and Barber (2010: 16) offer an even more conservative trend than Hanushek and Woessman. On the basis of trends observed in 12 education system a good and attainable trend is said to be 11.5 PISA[51] points in ten years, or 0.01 PISA standard deviations per year. The reason why this improvement level is so low is that ten of the twelve countries are developed countries. In the analysis that follows both developed and developing countries are analysed within an integrated model and one finding, as one might expect, is that there are diminishing improvements as countries reach higher levels of educational quality.

Educational quality targets, like other policy targets, are susceptible to what Inbar (1996) refers to as ritualistic planning, or planning where idealistic intention, rather than feasibility, becomes the overriding consideration when targets are set. A better empirical basis for educational quality targets can have the effect of reducing ritualistic target-setting, or at least clarifying that ritualistic targets are extremely unlikely to be attained. Yet, as Inbar (1996) would argue, making the target-setting process entirely scientific is often not possible as political considerations do weigh strongly.

[51] Programme for International Student Assessment.

How should targets involving a percentage of pupils be understood? Is there a simple way of translating such targets to improvements in standard deviations? Standard statistical methods do in fact allow for an approximate translation. To illustrate, one could consider a baseline of 20% of pupils attaining a minimum level of performance and the target that 60% of pupils should be able to attain the minimum standard. If one uses the Excel formula appearing as equation (68) below, *a* becomes 20% and *b* becomes 60%. The result is 1.09 standard deviations, or 0.27 standard deviations per year if one considers the four years for the South African target mentioned above. Clearly moving from 10% to 60% in four years is extremely ambitious. Moving from 50% to 60% in four years on the other hand translates to 0.06 standard devations per year and is, as will be confirmed below, feasible though ambitious. The Hawaii intention mentioned above to move from 10% to 100% in 13 years implies an improvement of 0.28 standard deviations a year, a virtually impossible target. The Excel NORMSINV function seen in equation (68) provides the *Z* score, or number of standard deviations, associated with a particular area under the normal curve. In the absence of Excel, the solution is best found using a standard table of areas under the normal curve (Blalock, 1979: 602).

$$Z = \text{NORMSINV}(1-a) - \text{NORMSINV}(1-b) \tag{68}$$

The analysis that follows will extract the best actual improvements over time to provide an indication of improvements that optimistic yet realistic planners could consider. However, before best trends are examined, it is instructive to examine the average trends seen in recent years. Table 3 in an earlier section summarised country-level mean test scores from recent years that were used for the analysis. That table indicated that for the period 2000 to 2009, 230 country-specific time series were available, where these consisted of two, three or four points not shared with any other time series. Of the 230 time series, 90 had two points, 21 had three points and 69 had four points. The mean annual improvement across all 230 series, with least squared trendlines used for the three- and four-point series, was 0.8 PISA points per year, or around 0.01 of a PISA-wide standard deviation. All country average values were converted to the PISA scale for this analysis using the method described in section 3.3. The average improvement was thus small, though positive. The figure becomes minus 0.002 points if only OECD countries are considered and 0.03 for non-OECD countries, suggesting that there is educational quality convergence between developing and developed countries. However, if the diminishing marginal rate of improvement is taken into account a simulation would indicate that it would take exceedingly long for developing countries to catch up. South Africa, for instance, would require 160 years to reach the levels found in non-Asian OECD countries. For Brazil the duration would be around 130 years. Clearly the mean rates of improvement seen in the data are discouragingly low.

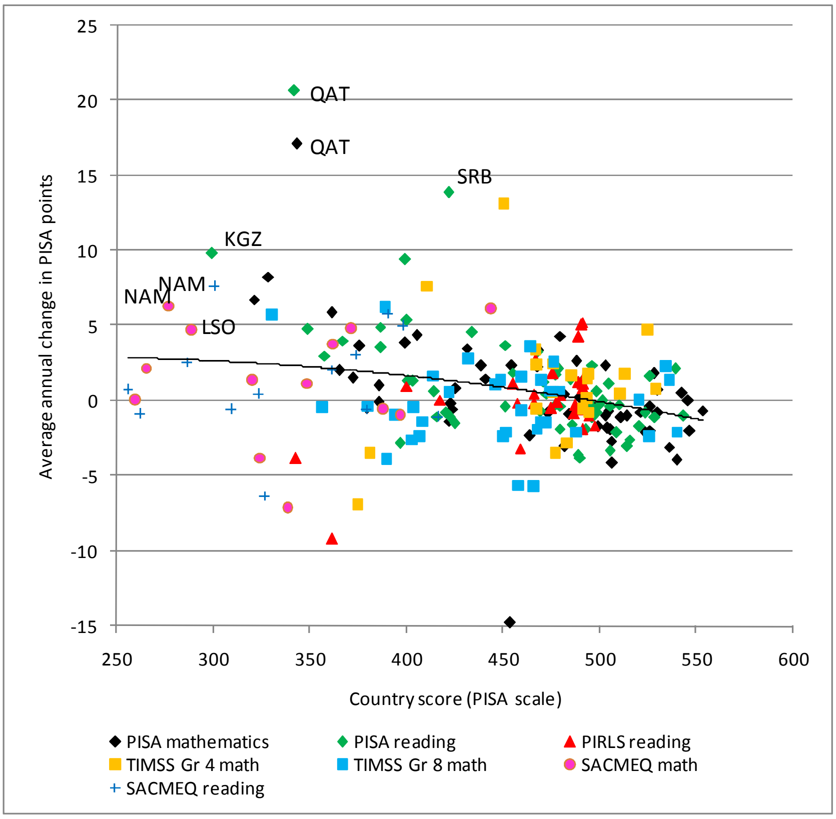
An analysis of the 90 series with more than two points suggests that there is considerable measurement error in the country scores. Of the 90 series, 15 display movements that are consistently upward, 15 display movements that are consistently downward and 60 display inconsistent trends consisting of both upward and downward movements. One might expect that of the 60, many would display small inconsistencies reflecting a situation of virtually no change. To test this, a margin of tolerance of 0.05 of the standard deviation of country scores per programme was used to determine when there was no change. If the change between two data points was within the margin of tolerance, the trend was considered a flat one with zero change. Using this margin still resulted in 47 series that displayed inconsistent trends moving both up and down. It seems unlikely that the quality of human capital would in reality display trends as erratic as what these figures suggest and reasonable to assume that the patterns seen in the data indicate measurement error. This measurement error should caution us above all against reading too much into the two-point series of individual countries. In fact, although these two-point series are used in the analysis that follows, the analysis was also done without these series. The exclusion of the two-point series did not make a difference that would substantially change the conclusions presented below.

The 15 consistently upward trends referred to above are found in 11 countries. None of the 11 countries had any series, from the set of 90 series, with a consistently downward trend. Examining the education policies of these 11 countries should be of particular interest. The 11 countries are, in descending order of size of their annual improvement: Brazil, Chile, Turkey, Indonesia, Tunisia, Poland, Germany, Israel, Hungary, Uruguay and Switzerland. The countries Chile, Tunisia, Turkey and Germany each had two series of more than two points with a consistently upward trend (for these four countries the average improvement across the two series was used for the ranking provided here).

The following graph displays the point increases of each of the 230 series referred to earlier. What immediately stands out is the greater degree of vertical dispersion amongst countries with lower country scores. This pattern remains if two-point series are removed. Part of the explanation for the greater vertical dispersion on the left-hand side of graph is that in the case of the 78 countries with more than one series (for instance because they have different series for different subjects in the same programme), within a single country trends are more likely to vary from series to series if the country performs poorly. Specifically, the within-country gap between the worst improvement and the best improvement amongst the 78 countries in question was on average 4.2 PISA points if the lowest half of performers were considered and 2.6 PISA points if the top half of performers were considered. Possible explanations are a greater degree of measurement error amongst developing countries and that developing countries, though they may on average display greater annual improvements, improve inconsistently and in spurts, perhaps due to policy or leadership instability.

The quadratic trendline in the graph confirms the pattern, albeit weak (R^2 is only 0.07), of greater increases amongst developing countries.

Figure 22: Annual country score increases by level



Sources: See Table 3.
Note: In order to position each point, which represents a time series, along the horizontal axis, original country scores had to be converted to the PISA scale. This was done by using the five sets of coefficients obtained through the process described in section 3.3. For each series, the mean of the converted country score values was used to plot the point on the horizontal.

Figure 22 is able to provide policymakers with a rough idea of the maximum possible improvements that should inform system-wide targets. Clearly annual improvements of as high as 0.2 standard deviations (or 20 PISA points) per year appear to be possible, though rare. At the same time it is important to acknowledge that of the 11 consistent improvers referred to earlier, the best average annual improvements within a single series were only 5.9 PISA points, in the case of Brazil, followed by 4.5 PISA points, in the case of Chile (these figures translate to 0.06 and 0.05 standard deviations).

Stochastic frontier analysis is conceivably an appropriate technique for identifying a systematic efficiency frontier, or a feasible outer limit for expected improvements in the data. However, the dataset being considered here is too limited in terms of the number of efficient units, or countries, for this technique. Instead a quantile regression approach was employed to find annual improvement levels that were amongst the best at various points along the horizontal axis of Figure 22. The three equations appearing below describe the method. These equations, or permutations of them, were used not just to identify the best educational quality trends but also the best trends with respect to other variables, for instance the best increases in life expectancy, in designing the growth simulation tool described below. Equation (69) refers to the OLS estimation of annual improvements β in the case of each three- and four-point series i discussed above. *Y* is the year of the test and *X* is the country test score at each point in time *t* within the series. A separate regression is run for each series. Next, equation (70) is used for a single quantile regression where the best expected annual rates of change in *X* across the various country-specific series *i* (the left-hand side of the equation) is predicted by the level of *X,* specifically the mean of *X* within each country-specific series*,* and the parameters *k*, *m* and *n*. In the case of educational quality, the quantile regression was run at the 90th percentile. Finally, equation (71) describes the derivation of what we can call the 'improvement curve'. The horizontal axis for this curve, when graphed, is time *t*, measured in years, and starting with year 0. The left-hand origin of the curve would reflect, along the vertical axis, the worst value *X* seen in the data, for instance the worst average country score. Subsequent points along the curve would be plotted using equation (71), up to the point representing what one would consider the best attainable value of *X*. Equation (70) allows for diminishing improvements over time, which is what one would expect. For most of the other explanatory variables used in the growth simulation the application of the three equations is simpler than in the case of educational quality, because one is dealing with just one time series per country, so *i* is simply a country observation. What makes the educational quality data unusual is that we have data for only a short historical period and that the data are derived from several different testing programmes.

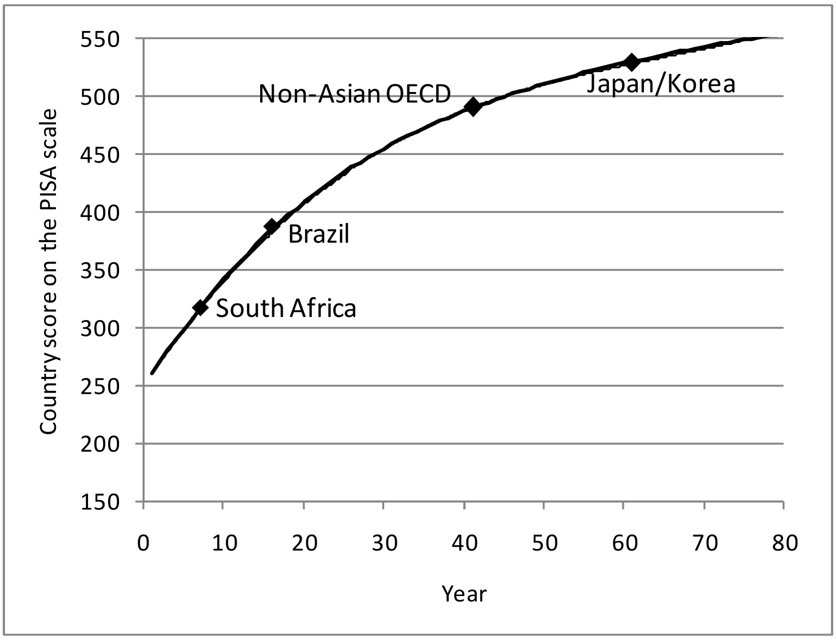
$$X_{it} = \alpha_i + \beta_i Y_{it} \tag{69}$$

$$\left(\frac{\beta}{\overline{X}}\right)_i = k + m\overline{X}_i + n\left(\overline{X}_i\right)^2 \tag{70}$$

$$X_t = X_{t-1} \times \left(k + mX_{t-1} + n\left(X_{t-1}\right)^2\right) \tag{71}$$

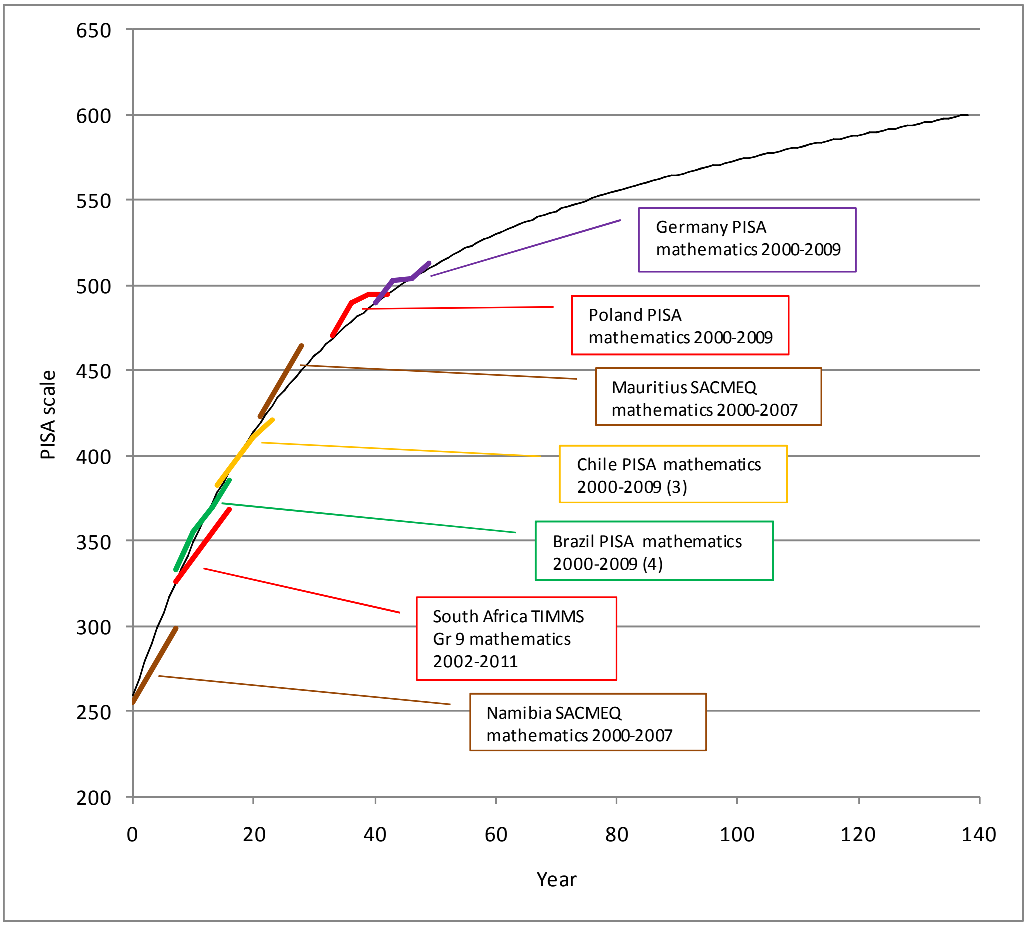
The curve in the following graph is obtained using equation (71). Pseudo R^2 in the quantile regression was 0.23. The graph illustrates, for instance, that even if it improved at more or less the best possible rates seen around the world, South Africa would take 34 years to attain the level of performance of non-Asian OECD countries (average of 491 on the PISA scale). It would take Brazil about 25 years to achieve the same. These are clearly optimistic scenarios, but not impossible ones. One could translate the scores reflected on the vertical axis to the metric of another testing programme using the coefficients from equation (33).

Figure 23: Minimum years needed to improve



The following graph expands on the previous one by superimposing actual improvements from key improver countries. Here South Africa is included, on the basis of trends revealed after the release of TIMSS 2011 grade 9 results, in 2012. TIMSS 2011 values were not included in the calculations described in section 3.3. A graph such as this one brings together a number of issues policymakers would be concerned about: level of performance, speed of past improvements and feasible future improvements.

Figure 24: Minimum years needed to improve with actual examples



Sources: See Table 3, also Mullis, Martin, Foy and Arora (2012).
Note: Numbers in brackets refer to the number of data points available, where more this is more than one point.

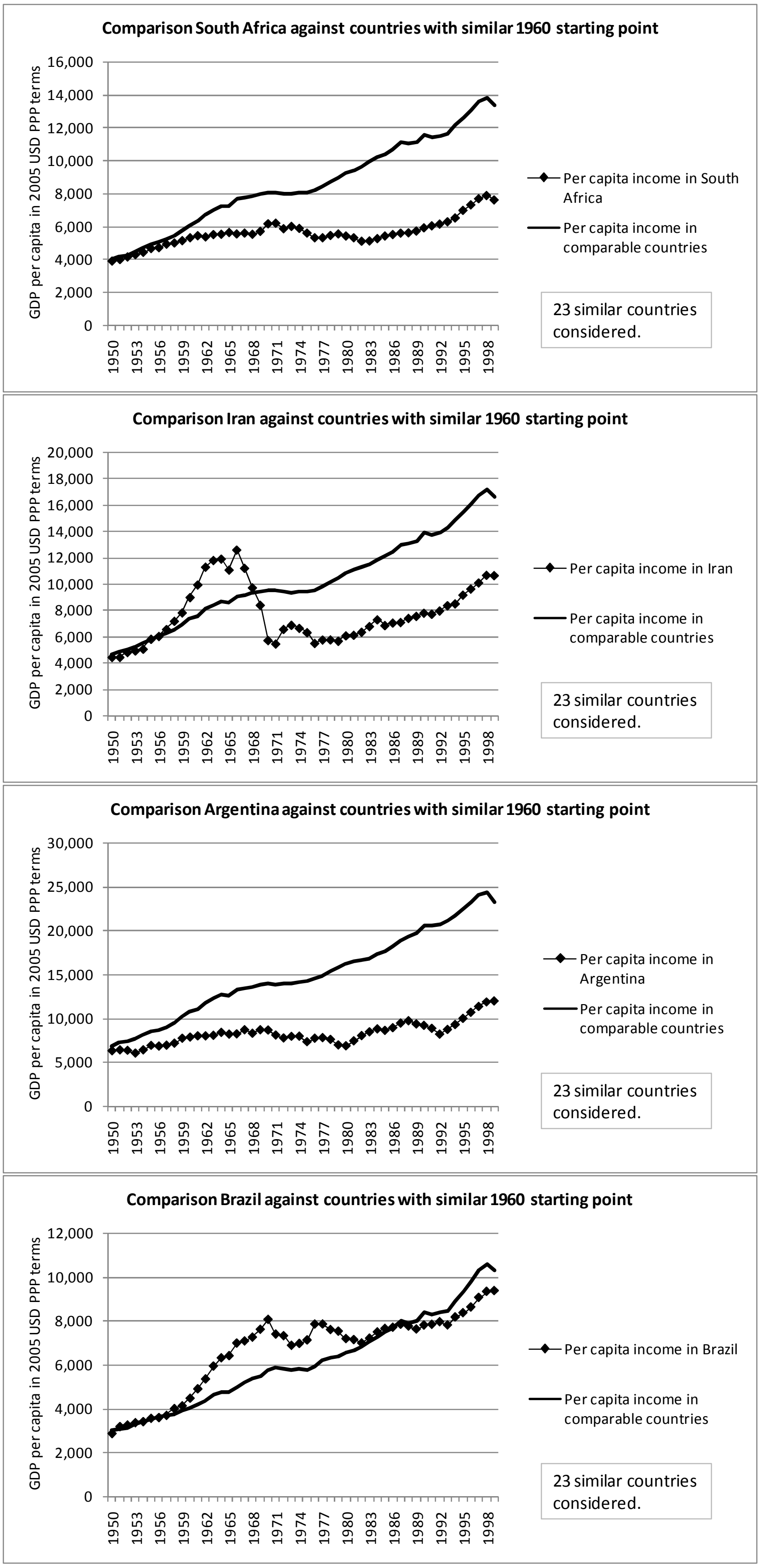
5.4 Magnitude of educational and non-education under-performance

There is much emphasis in the development literature aimed at policymakers on comparing key development indicators across countries. For instance, the *World Development Report* of the World Bank (for instance 2011b) presents a selection of World Development Indicators where countries can be compared across the same indicator. UNESCO's (for instance 2011a) annual *Global Monitoring Report* does something similar for just education, though it is noteworthy that no learning outcomes values have been presented as indicators along the lines of what is seen in Table 21 in the appendices. The World Economic Forum (for instance 2012), in its *Global Competitiveness Report*, presents country indicators in a somewhat different fashion, using country as a principal category and within that providing rankings for the country in question. Given the degree to which these kinds of reports seem to stimulate public debate, as seen in Matola (2012) with respect to the 2012/13 *Global Competitiveness Report*, it seems they succeed in providing one bridge between more academic analysis and the general policy discourse. Yet these reports have their drawbacks. Above all, in the attempt to compile large amounts of comparable statistics, scrutiny of data quality is compromised, resulting in a situation where very often statistics are not really comparable across countries. To illustrate, UNESCO indicators that draw from enrolment and population data to produce gross enrolment ratios have been criticised because they do not sufficiently adjust for or even acknowledge problems of comparability between the two data sources. This is argued in Gustafsson (2012b). A further problem is that statistics may be poorly explained. The World Education Forum (WEF) educational quality indicators are widely understood to represent scientific measures of what pupils actually learn. Such an interpretation is evident in Matola (2012). However, closer examination of the WEF's methodology notes reveals that educational quality values are drawn from a few country-based experts who evaluate the education system on a seven-point scale. An expert opinion approach is thus employed, but in a manner which in no way is oriented towards cross-country comparison. It is therefore not surprising that, for instance, South Africa's primary schooling should emerge as of a poorer quality than that of Lesotho, when figures such as those in Table 21 suggest that the opposite is true. Moreover, South Africa emerges as worse than several countries, around 40, which have never participated in any international testing programme.

It is not easy to prescribe solutions for these problems in the international indicators. The demand for these indicators is high, yet because policy decisions based on them are fluid rather than hard in the sense that they do not, for instance, influence the distribution of global resources in any precise manner, users are often prepared to overlook errors, sometimes serious ones. It would be difficult, even for organisations such as UNESCO and the World Bank, to invest sufficiently to achieve reliable statistics. This could require running new surveys in countries where data are known to be unreliable. However, what would be feasible is better documentation relating to what countries and what indicators are most problematic, based on internal, including historical, consistency checks. Such documentation is currently almost non-existent.

Section 5 includes some illustration of the relative performance of single countries with respect to key development indicators. Specifically, apart from educational quality, seven non-education indicators are examined, partly so that the magnitude of the educational quality challenge can be viewed in the context of other development challenges. But before explanatory variables are examined, it is worth looking at trends with respect to the dependent variable, economic growth. The four countries represented in Figure 25 demonstrate how precarious growth can be and specifically how political circumstances can compromise growth and average income. In discussing growth, it is always worth considering how costly the risks are associated with political instability and ideology that is divorced from basic notions of well-being. South Africa's unfortunate political history, in which the human capital of black citizens was curtailed in various ways, has resulted in relatively weak growth since at least 1960. South Africa's average income figure in 2009 was around 57% of what was seen in other countries that started off at more or less the same income level as South Africa in 1960. The graph is taken from the Excel growth simulation tool and the 23 comparator countries are countries in the decile immediately above and below South Africa, meaning more or less a quintile of countries with South Africa as the median, using the per capita income in the starting year, 1960 in this case, as the ranking variable. Iran saw a calamitous drop in average income that coincided with the Iranian Revolution, after an initial couple of decades where growth exceeded that in the comparator countries. In 2009, Iranians, like South Africans, were experiencing a level of income well below what they presumably would have enjoyed had their political history been different. Argentina and Brazil, in many respects similar but politically rather different, offer an interesting comparison. In 1960 Argentina's per capita income was over double that of Brazil, but political instability, populist governments and economic mismanagement took a heavy toll and by 2009 Argentina's income was only 25% greater than Brazil's. In fact, many of Argentina's initial comparator countries, such as Spain, had become classified as high income countries by 2009. Brazil, despite experiencing political instability similar to that of Argentina, succeeded in maintaining a relatively stable economic climate that promoted investment in appropriate human and physical capital.

Figure 25: Growth trajectories in four countries



For the growth simulator tool, variables were selected which, according to the literature, were significant predictors of growth and which were amenable to policy interventions. Educational quality is clearly one such variable. An example of a variable that predicts growth relatively well but is not subject to policy interventions would be the proportion of the country situated within the tropics. Even the proportion of the population that is Confucian, a variable found to be significant in Sala-i-Martin, Doppelhofer and Miller (2004) (or SDM), can be considered beyond the scope of normal policy interventions. Moreover, it had to be possible to obtain data on the variable from a recent year that could serve as a baseline, given that the aim was to use analysis already conducted of past trends to forecast trends into the future. In the tool, any year beyond 2009 was considered a future year. Measurement difficulties, though present, were never large enough to exclude any variable, meaning that ultimately significance and amenability to policy-induced shifts were the determining criteria in the selection of the eight variables, one of which was educational quality.

Of the seven explanatory variables other than educational quality, three are variables that were found to be prominent predictors of growth in SDM and that were re-used in the growth simulator tool with no adaptations. The three variables are: investment price, life expectancy, and government consumption as a proportion of GDP (see Table 10). Three explanatory variables, malaria prevalence, trade openness and ethnolinguistic fractionalisation, are also amongst the most prominent SDM variables, but because it was not possible to replicate measurement in the way SDM did, different ways of measuring the baseline values in the growth simulator tool had to be found (the details of this are explained below). For these three variables, adjustments also had to be made to the slope coefficients in the SDM growth regression. Finally, one variable that draws from the growth modelling of Hanushek and Woessman (2009) is property rights, which is referred to as rule of law for the purposes of the tool (see for instance Table 16).

The next table summarises the baseline values from a recent year with respect to all eight explanatory variables. The year for the baseline values was 2009, or a year as close as possible to 2009. For certain variables, the baseline year was different for different countries. The tool explains what year applies for each variable with respect to the country that is selected. The maximum possible number of countries is 152, which is the number of countries for which per capita income values existed in Heston, Summers and Aten (2011). In keeping with analysis found in the literature, very small countries, in the case of the tool countries with a population of less than one million, were excluded (see justification in section 4.2).

Table 19: Details on explanatory variables for growth

Variable	*Source*	*N*	*Mean*	*Std. dev.*	*Min.*	*Max.*
Educational quality	See section 3.3.	100	425	66	260	578
Investment price	Heston, Summers and Aten, 2011.	150	67.3	28.7	0.0	180.0
Malaria prevalence	Online data-querying facility of World Bank (World Bank DataBank), consulted 2012.	152	0.0282	0.0652	0.0000	0.3477
Life expectancy	Online data-querying facility of United Nations Statistics Division (World population prospects: The 2008 revision), consulted 2009.	151	69	10	45	85
Trade openness	Heston, Summers and Aten, 2011.	149	83	50	2	382
Ethnolinguistic fractionalisation	Alesina, Devleeschauwer, Easterly and Kurlat, 2003.	147	0.406	0.288	0.002	0.923
Government consumption share	Heston, Summers and Aten, 2011.	149	11.1	6.9	3.0	62.0
Rule of law	World Bank, 2010.	152	-0.16	1.01	-2.43	1.97

What follows is some discussion of the values used for each of the eight variables and the relevance the variables may have for education planning. How feasible improvement trends were identified for the last seven variables, along the lines of the analysis of educational quality in section 5.3, is discussed in section 5.5

Educational quality. This variable was discussed extensively in section 3.3. Of the 113 countries for which values are found in Table 21, 100 were found in the set of 152 countries with the required income data and thus included in the tool.

Investment price. As in SDM, for the tool the Penn World Table (Heston, Summers and Aten, 2011) is the data source, though for the tool a later version of this source is used. Values reflect, in terms of USD PPP, the prices of inputs needed in construction as well as machinery and equipment needed in industrial processes. A large range of inputs, from vehicles to computer software, are considered. Because values are expressed in PPP terms, they would not reflect the attractiveness of investment for foreign investors, but rather for local investors. A country's tax system is one factor that might change its PPP investment price. The same PPP indices used for GDP in general are used to obtain PPP prices for just investment inputs. To illustrate, particularly highest investment price values are found in Cuba, Norway and the Republic of Congo. Countries with low values include Botswana, Saudi Arabia and Vietnam. Prices in developed countries tend to be higher. Of interest to the education planner would be the fact that if skills needed for investment activities, such as engineering skills, are in particularly short supply, the PPP investment price would rise and in general growth can be expected to be slower.

Malaria prevalence. It was not clear how the SDM values for malaria prevalence were obtained. The source that is referred to in SDM (2004: 820), Gallup, Mellinger and Sachs (2001), contains values that are substantially different to those of the dataset that accompanies SDM and no way could be found of deducing what type of adjustment occurred when the values used by SDM were prepared. For the tool, malaria prevalence values from the World Bank's DataBank were used. These data have their own problems, specifically erratic trends for certain countries. For instance, Tanzania moves from one million to 30 million to half a million malaria sufferers between 1997 and 1999, a trend which must be an error. According to Barat (2006:

15), in countries that have succeeded in reducing malaria prevalence, the education authorities have played an important role.

Life expectancy. The UN data on life expectancy display a high degree of internal consistency. The education system influences life expectancy, and vice versa, in a number of ways. For instance, education may promote behaviour that reduces HIV infections, which in turn increases life expectancy. On the other hand, higher average life expectancy means that human capital, once developed, is available for a longer period of time, which can be beneficial for the inter-generational transfer of knowledge and skills.

Trade openness. The data used by SDM, which referred to years during which the economy had been open, could not be found. As a substitute, Penn World Table values referring to imports plus experts as a percentage of GDP were used. This indicator is used by, amongst others, Barro and Sala-i-Martin (2003: 529) for trade openness.

Ethnolinguistic fractionalisation. This is a variable with an interesting history which is of special relevance for education planners, given the importance of education systems in promoting international and within-country communication through the teaching of common languages. In Sub-Saharan African countries such as South Africa, linguistic diversity produces special challenges in the development process. The same can be said of Asian countries such as India. This seems to justify paying special attention to this variable. In section 5.5, a new approach for an ethnolinguistic fractionalisation index is proposed, which is arguably more sound conceptually than approaches that are widely used, though the proposed approach may be difficult to implement across a sufficiently broad range of countries due to data limitations. The proposal draws from an unusual South African data source.

SDM employ a commonly used set of values that originates in an atlas published in the Soviet Union in 1964. Each value reflects the 'probability that two randomly selected people from a given country will not belong to the same ethnolinguistic group' (Easterly and Levine, 1997: 1243). Values are based mainly on the languages people speak, though to a limited extent other ethnic factors influence the index too (Alesina, Devleeschauwer, Easterly and Kurlat, 2003: 156). As pointed out by Fearon (2003), how one defines language and ethnicity has a large impact on the relative degree of fractionalisation found in specific countries. An advantage with the Alesina *et al* (2003) values over values drawn from the Soviet atlas is that the former source is more explicit with respect to definitions. In fact, Alesina *et al* (2003) base their composite index on three separate indices dealing with language, religion and a 'broad measure of ethnicity' (Alesina *et al,* 2003: 156). Moreover, this source uses relatively recent data, for instance language data from 2001. In both sources, the index for one country is calculated as one minus the Herfindahl index, an index traditionally used to gauge the degree to which an industry is concentrated in just a few firms.

$$F = 1 - \sum_{m=1}^{n} p_m^2 \qquad (72)$$

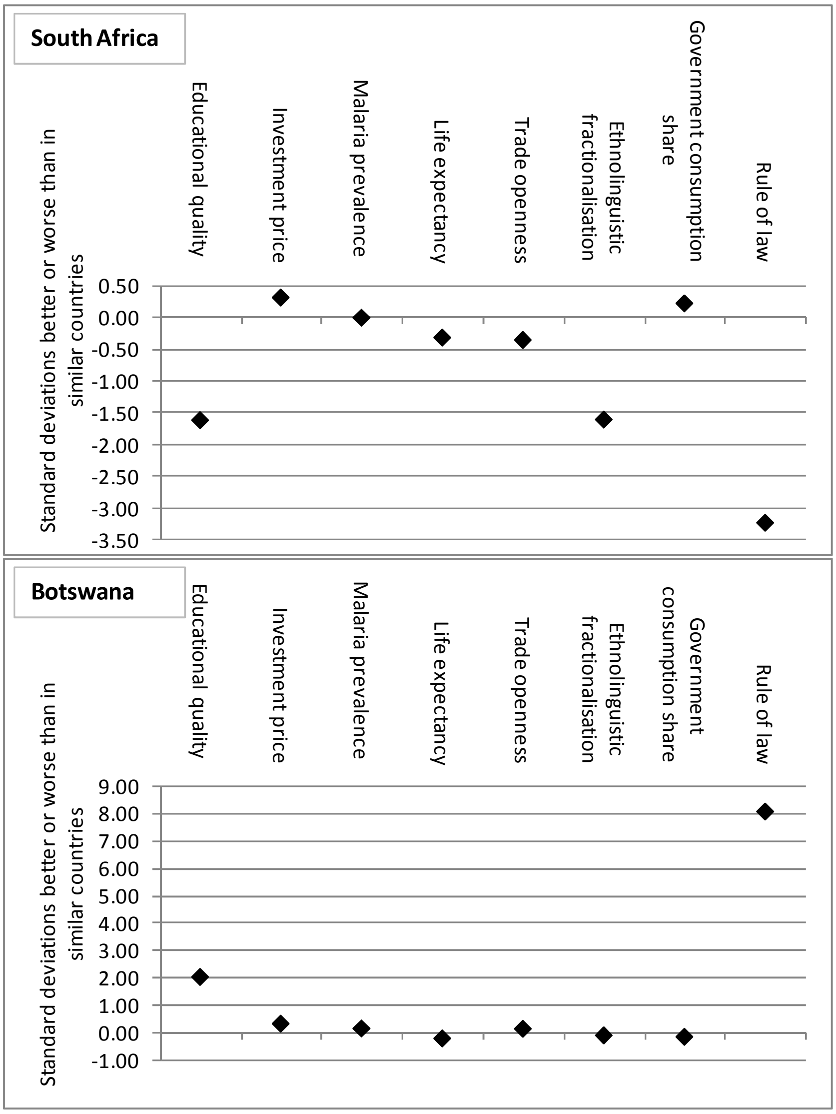
The fractionalisation index value for country *j* is thus calculated using *n* values of *p*, where *p* is the proportion of the population within each of the *n* groups according to, for instance, home language. The index value can range from 0.0, where everyone belongs to the same group, to 1.0, in the theoretical instance of a society where each individual belongs to his own group. To illustrate, the index values in Alesina *et al* (2003) for language are (with the Easterly and Levine values, which were based on the original Soviet atlas, in brackets): Japan 0.02 (0.01); United States 0.25 (0.50); India 0.81 (0.89); and South Africa 0.87 (0.88). The language-based index values from Alesina *et al* (2003) were used in the simulation tool as they were methodologically the closest to the original SDM values (both focussed strongly on language), they follow a scale which seems very close to the original SDM scale, and as they are relatively recent values (as mentioned earlier, all from 2001).

Government consumption share. This is government consumption as a percentage of GDP. The data source used for the tool is the Penn World Table. Though SDM use a different source, the 2009 mean from Penn of 11.1 is close to SDM's mean of 11.6.

Rule of law. As in the previous section 4.4, the World Bank's Worldwide Governance Indicator is used, though for the tool a 2011 update of the data were chosen, which included values for years in the range 1996 to 2010. For the tool, 2010 values were used for the baseline.

The next two graphs offer an illustration of the relative under-performance of individual countries with respect to the eight explanatory variables described above. In the case of four variables, a lower value is considered better for growth: investment price, malaria prevalence, ethnolinguistic fractionalisation, government consumption share. The graphs for South Africa and Botswana, both produced within the tool, are instructive. Again, similar countries are the countries in the decile above and the decile below the country in question, using present income per capita as the sorting variable. South Africa's values are better than the average across similar countries for just two variables, investment price and government consumption share, and the difference is small, less than 0.5 of the standard deviation across all 152 countries. In the case of three explanatory variables, educational quality, ethnolinguistic fractionalisation and rule of law, South Africa performs substantially worse than countries with similar incomes. The situation in neighbouring Botswana is very different. Botswana does not under-perform to a large degree with respect to any of the eight explanatory variables. One should keep in mind that these eight variables are variables that have been found to be particularly important for growth. Botswana performs exceptionally well in terms of one variable, rule of law. The graphs serve to alert the South African planner that certain development areas require more urgent attention than others, but also that ideas for the needed policy solutions might be found in countries such as Botswana.

Figure 26: Relative sizes of development challenges



One thing the growth simulation tool does not consider explicitly is the impact of educational quantity, in the sense of the proportion of the young population enrolled in schools or average years of schooling amongst adults, on growth. The work by Hanushek and Woessmen (2009) has suggested that educational quality is a far better predictor of growth than educational quantity and this seems to justify having a simulation that excludes an explicit treatment of quantity. A further justification for this approach is that there does not appear to be a trade-off between quality and quantity in education of the kind one may intuitively expect. Put differently, increasing the proportion of children enrolled, which generally means improving the coverage of children from the poorest households, seems not to be associated with a decline in the average test scores of pupils. Improving quality versus increasing enrolments would therefore not appear to be a choice that policymakers have to make. Both indicators can improve simultaneously. This is demonstrated in Taylor and Spaull (2013) using SACMEQ[52] data. The absence of a clear quality-quantity trade-

[52] Southern and Eastern Africa Consortium for Monitoring Educational Quality.

off, coupled to evidence that quality predicts growth to a far greater degree than quantity, seems to justify the exclusion of enrolment from the growth simulation tool.

5.5 The relative magnitude of education's growth effect

In various texts Hanushek and Woessman have taken the outputs from their growth regressions and used them to project likely changes to future growth and national income if certain increases to average test scores are realised (Hanushek and Woessman, 2007; Hanushek and Wößmann, 2007; OECD, 2010a). Their most comprehensively explained forecast is arguably that of OECD (2010a), titled *The high cost of low educational performance*. In this text the focus is on OECD countries only. The authors argue that policymakers and governments do not pay enough attention to educational quality improvements, partly because they are not aware of how large the future welfare benefits of such improvements are, but also because of a general absence of long-term planning and an over-reliance on more short-term growth stimulation, in particular demand-focussed short-run business cycle management. Put differently, there is an under-emphasis of supply side interventions such as growing human capital. In OECD (2010a) Hanushek and Woessman draw from Lucas's (2003: 1) presidential address to the American Economic Association, which though not dealing specifically with education or human capital, emphasises the importance of a better focus on long-range growth dynamics:

> Taking performance in the United States over the past 50 years as a benchmark, the potential for welfare gains from better long run, supply-side policies exceeds *by far* the potential from further improvements in short-run demand management.

In OECD (2010a) three future scenarios are presented. In the first, OECD countries each improve their average PISA score between 2010 and 2030 by 25 PISA points, or a quarter of a standard deviation in PISA (see the definitions in section 3.3). In the second scenario, all countries move to the level of the highest baseline performer, Finland. In the third scenario, the within-country distribution of scores is adjusted so that eventually no pupil has a score under 400 PISA points. The first two scenarios, and perhaps the third scenario, though this is not completely clear, involve improvements in the performance of pupils during the period 2010 to 2030. By 2071, the target level of human capital is assumed to exist throughout the labour market, across all ages, so starting from this year the difference between the baseline growth rate, which is based on recent historical data, and the steady-state growth rate is the slope coefficient for the educational quality variable in a growth regression such as that of model (3) in Table 14. They refer to the baseline growth rate as the 'education as usual' rate (OECD, 2010a: 22). The coefficient value used is 1.7, in the sense of a 1.7 percentage point increase in the growth rate for every 100 additional PISA points. The value is obtained from a model based on the data for just 23 OECD countries, where the model includes initial average years of schooling amongst adults and initial per capita income. Years of schooling carries a statistically insignificant coefficient in the model. Equation (73) expresses the steady-state calculation: the difference in the annual growth rate g is the slope coefficient B multiplied by the improvement in educational performance T. Before 2071, additional growth is smaller than in equation (73) for two reasons. Firstly, it takes 20 years for the PISA improvement to be brought about in schools and even then, by 2030, it takes a further 40 years before all adults in the labour force have the new level of human capital.

$$\Delta g_Y = B \times \Delta T \tag{73}$$

In the third scenario, the underlying growth regression is more complex, with two explanatory variables drawn from the performance data, namely the proportion of pupils achieving less than 400 (this would carry a negative coefficient) and the proportion achieving more than 600 (a positive coefficient).

Results of the forecasting are expressed per country in three ways: future GDP as a percentage of baseline GDP; the total of the 2010 to 2090 stream of additional GDP expressed as a present value, using a discount rate of 3.0%; and the long-run additional percentage points to annual GDP growth, relative to the baseline. The third statistic might appeal to policymakers concerned with politically set growth targets, such as South Africa's target of 5% (set in 2012) or China's target for the 2011 to 2015 five-year period of 7% (South Africa: National Planning Commission, 2012: 39; KPMG, 2011). The first statistic, future income as a percentage of baseline income, has the advantage that it reflects the gains obtained through the compounding of many years of growth, a gain that often appears counter-intuitively high, meaning intuition tends to under-value this gain. The present value of future income streams is a convenient basis for comparing education interventions to other interventions and of dealing with the complexity brought about by the lags in the gains associated with better education.

OECD (2010a) includes margins of error for forecasted gains in GDP based on a confidence interval in the slope coefficient that is based on a 5% level of significance. Given the effects of compounding, these margins are large. For instance, for all 23 OECD countries considered in the first scenario, total GDP in 2110 is expected to be 33% higher than in the business as usual scenario, though the margin for this improvement is 15% to almost 60%. The authors present a number of fairly obvious caveats relating to the forecast. One is that the growth regression informing the coefficients used for the forecast could be inaccurate. In this regard, it is worth re-stating the point made in section 4 that the size of the education quality coefficient is highly sensitive to the overall number of variables in the regression. Due to multicollinearity, the greater the number of variables, the smaller the coefficient. This was seen in Hanushek and Woessman (2009) and was also evident in Table 14 and Table 16 above, which indicated that a more inclusive model reduces the size of the educational quality coefficient by two-thirds. It could be argued that the coefficient used in OECD (2010a) by Hanushek and Woessman is three times as large as it should be. A further caveat acknowledged more explicitly in OECD (2010a) is that there is a risk in assuming that growth over the next 100 or so years will occur within the same structural parameters as that over the last 50 or so years. As discussed in sections 3.4 and 3.5 above, there is considerable debate over defining and measuring growth, partly because economic growth acquires different meanings in different historical settings. One parameter that is certain to change is the predictability of climate. As argued in the *Stern Review*, climate change can be expected to reduce growth by one percentage point by 2050 relative to a situation with no climate change (see section 3.5). Hanushek and Woessman do in fact make use of the *Stern Review* when they decide on a discount rate to use (OECD, 2010a: 22).

A forecasting exercise with a similar policy focus but a rather different method to that of Hanushek and Woessman in OECD (2010a) is that of Patrinos and Psacharopoulos

(2011). Comparing the two texts is instructive. The latter text, like the first one, presents estimates of forfeited income growth associated with a failure by governments to invest sufficiently in education. The difference is that Patrinos and Psacharopoulos focus not on educational quality but quantity, in particular the average years of education attained amongst adults in a country. Why they chose to focus on quantity, rather than quality, is not made explicit but it appears they were interested in obtaining as many data points as possible, in terms of countries and historical time periods, and this would preclude an analysis based on average test scores. Their historical period of interest is long, from 1900 to the present. They do acknowledge the possible tenuousness of years of schooling as an indicator of actual human capital, or of the knowledge and skills people possess. Much of the modelling makes use of a 'macro-Mincerian function' with just two variables: the log of average country income on the left and the average years of schooling on the right (Patrinos and Psacharopoulos, 2011: 20). A total of 1,596 observations are used, each being specific to a country and a year in the data. They include no explanatory variables other than years of schooling, a decision that is likely to inflate the years of schooling coefficient as the effect of institutional factors such as promoting the rule of law would have been excluded. Like Hanushek and Woessman, Patrinos and Psacharopoulos (2011: 19) find that that forfeited income is large: for instance, if during the period 1900 to 2050 Pakistan had an additional year of schooling across all years, income would have been around 20% to 40% higher in all years. Both texts appear convincing for policymakers. Clearly, policy conversations based on the relationship between years of schooling and growth cannot be regarded as redundant. The conversation is inevitably constrained by the data that are available at any point in time.

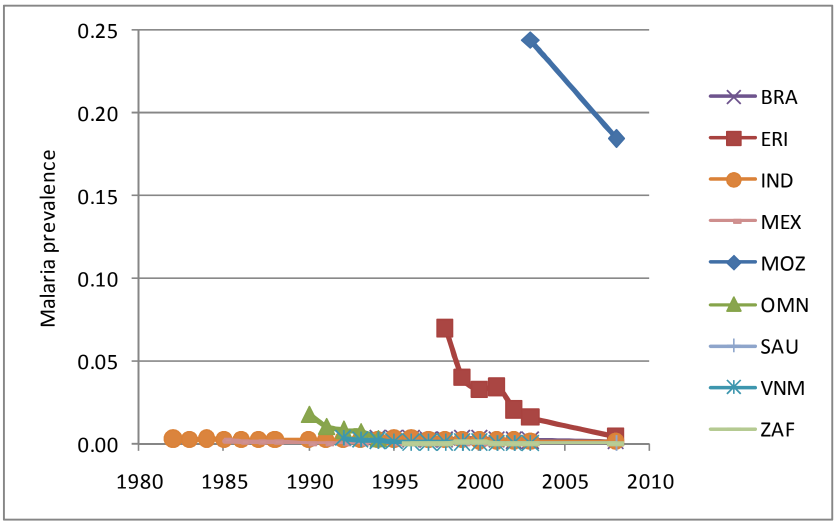
In attempting to forecast income gains associated with particular education quality improvements, the growth simulation tool expands on the work found in OECD (2010a) in three key respects. Firstly, a lower slope coefficient for educational quality is used, in recognition of the finding that this coefficient is lower than what is seen in OECD (2010a) if a larger range of variables is included in the growth regression. This seems appropriate as one can assume that to some degree policy interventions focussing on other variables are implementable independently of educational quality interventions. Other interventions of note would include health advocacy campaigns, language policies aimed at promoting social cohesion and initiatives aimed at instilling respect for the law.

Secondly, the forecasting in the tool includes explicit details on the assumed impact of variables other than educational quality as, notwithstanding the above point about the separateness of policy interventions, to some extent education policymakers regard themselves as exerting an influence on aspects of society beyond schooling. Thirdly, the tool, in drawing from the analysis of section 5.3, uses more empirically informed assumptions around the speed with with educational quality can be improved. But in the design of the tool, feasible speeds of improvement were also sought for the non-education variables. It was necessary to produce a curve for the non-education variables similar to the education improvement curve of Figure 23. The following paragraphs discuss how this was achieved for the seven explanatory variables in the tool other than educational quality. The seven resultant graphs are not reproduced here but can be found in the tool.

Investment price. For this variable, where a decline represents an improvement, equations (69) to (71) were used without any adjustments. Actual historical declines in the Penn World Table data were considered sufficient to permit an empirically informed improvement trend. A floor was set at 45 USD, or the 25th percentile, as it was assumed that moving to an even lower investment price would involve interventions that are not easy to replicate across countries, for instance highly generous public subsidies or lowering the unit cost of labour.

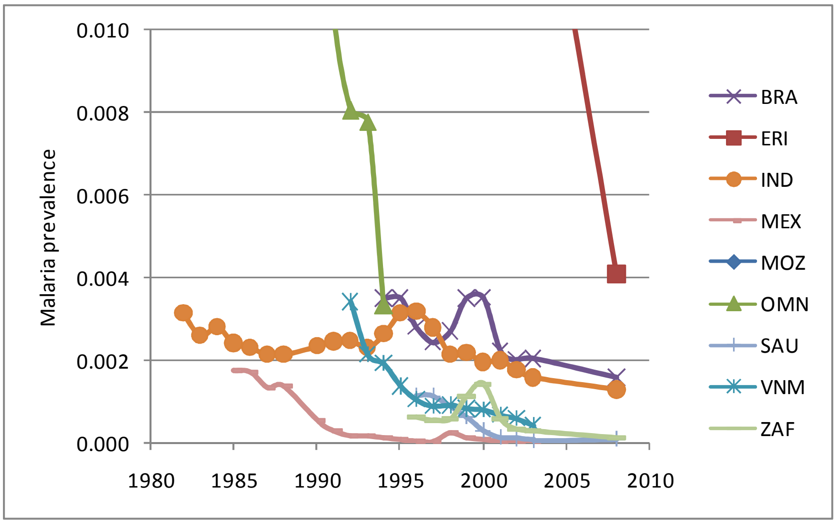
Malaria prevalence. Here an account by Barat (2006: 12) of eleven countries with exceptional improvements was used as a point of departure. Nine of the eleven countries displayed clear improvement trends in the World Bank data. The trends for the nine countries are displayed in the next two graphs, the second of which is a magnification of the first. The three equations referred to above were used with the data from the nine countries.

Figure 27: Malaria prevalence reduction I



Source: Online data-querying facility of World Bank (World Bank DataBank), consulted 2012.

Figure 28: Malaria prevalence reduction II



Source: Online data-querying facility of World Bank (World Bank DataBank), consulted 2012.

Life expectancy. The three equations were used with an assumed maximum life expectancy of 79 as the data suggested extremely slow improvements beyond this point.

Trade openness. Here improvements (increases) were only implemented if the value was below 100. The intention was to focus on improving trade openness up to a level just above the baseline mean in the case of more closed countries. The focus was thus not on seeing already fairly open countries become more open, something that is difficult to model without an explicit consideration of circumstance, such as individual country size.

Ethnolinguistic fractionalisation. Here it was not possible to use the three formulas as there was no trend data for the variable. As explained earlier, all values were from around 1964. As Fearon (2003) argues, it is unrealistic to assume that ethnolinguistic fractionalisation does not change over time. Roeder (2001) puts forward country values for 1961 and 1985 but it is not clear whether there is sufficient comparability across the two years as it seems possible that rather different methods were used.

Clearly the utilisation of languages evolves and this may lead to less fractionalisation. One just needs to the think of the rise of English as a world language in the last half a century. In an attempt to obtain some sense of the magnitude of the changes possible within one country, but also to explore an under-explored issue in South Africa, a rather unusual South African household dataset was used where not only the home language of respondents was collected (this is common) but also all the national languages other than the home language spoken by respondents (this is seldom collected). Arguably one needs data on all the languages spoken by individuals in order to model properly the degree to which people are divided from each other by language. Moreover, a picture of languages spoken other than home language is of particular importance to education policymakers as this picture can be changed relatively easily through education (the distribution of home languages is less alterable).

The South African dataset in question covers a nationally representative sample of around 5,700 households in 2004 (South Africa: HSRC, 2005) . Two key language questions were asked: ‘Which language does (the person) speak most often at home?’ and ‘What other languages can the person speak?’. Possible responses were limited to the eleven official languages and ‘Other’. ‘Other’ was selected by just 0.6% of respondents in the case of the first question. The following table illustrates the distribution of home and non-home language abilities.

Table 20: Language use in South Africa amongst adults (2004)

	Total home lang. (m)	Total speakers where not home language (millions)											Total speak-ers
		AFR	ENG	NBL	XHO	ZUL	NSO	SOT	TSN	SSW	VEN	TSO	
Afrikaans (AFR)	4.1		3.5	0.0	0.1	0.1	0.0	0.1	0.1	0.0	0.0	0.0	9.4
English (ENG)	2.7	1.7		0.0	0.1	0.2	0.0	0.1	0.0	0.0	0.0	0.0	20.2
isiNdebele (NBL)	0.6	0.2	0.4		0.0	0.3	0.2	0.0	0.1	0.0	0.0	0.0	1.2
isiXhosa (XHO)	4.7	0.5	3.5	0.0		0.6	0.0	0.4	0.1	0.0	0.0	0.0	6.1
isiZulu (ZUL)	7.4	0.4	4.4	0.2	0.6		0.2	1.1	0.2	0.2	0.0	0.2	11.3
Sepedi (NSO)	2.8	0.7	1.8	0.2	0.1	0.8		0.4	0.6	0.0	0.1	0.3	4.0
Sesotho (SOT)	1.9	0.6	0.9	0.0	0.2	0.6	0.1		0.3	0.0	0.0	0.0	4.5
Setswana (TSN)	2.4	0.9	1.5	0.1	0.2	0.4	0.2	0.3		0.1	0.0	0.2	4.0
siSwati (SSW)	0.9	0.2	0.6	0.0	0.0	0.6	0.1	0.1	0.1		0.0	0.1	1.4
Tshivenda (VEN)	0.6	0.1	0.4	0.0	0.0	0.1	0.2	0.0	0.0	0.0		0.1	0.8
Xitsonga (TSO)	0.8	0.2	0.5	0.0	0.0	0.2	0.2	0.2	0.2	0.1	0.0		1.8
Total	28.9	5.4	17.5	0.6	1.3	3.9	1.2	2.6	1.6	0.5	0.2	1.0	

Source: South Africa: HSRC (2005).
Note: Figures are in millions (sample weights used). Anyone aged 15 or above is considered an adult. Each value in the last column is the sum of two values: the number of home language speakers reflected in the first column and the number of speakers who do not use the language as a home language (this is reflected in the bottom row of the middle panel). Thus the 20.2 million English speakers comprise 2.7 million home language speakers plus 17.5 million who speak English as a non-home language. There are 28.9 million adults in total reflected in the table. Three-letter codes are ISO codes.

An index value using the South Africa data which is comparable to the SDM value would be the probability that two randomly selected adults would *not* have a language that they could speak to each other in. Put differently, the index would reflect the average percentage of other South Africans that each South African would not be able to communicate with. The calculation of the index is illustrated in the following three equations.

$$e = 1 - \frac{C}{T} \tag{74}$$

$$T = \frac{P \times (P-1)}{2} \tag{75}$$

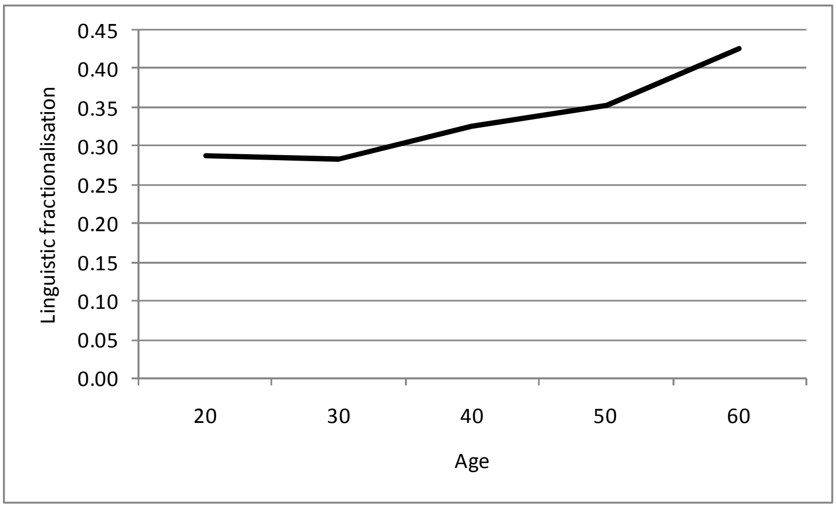
$$C = \left(\sum_{i=1}^{m} \frac{P_i \times (P_i - 1)}{2} \right) + \left(\sum_{i1=1}^{m-1} \sum_{i2=i1+1}^{m} d_{i1,i2} P_{i1} P_{i2} \right) \tag{76}$$

The linguistic fractionalisation index e proposed here is one minus a ratio which is the total number of two-way conversations possible amongst adults, given the languages people command, divided by the total number of two-way conversations which would be possible if everyone spoke just one language. The denominator T is a function of the total number of adults P according to equation (75). It is assumed that non-adults should be excluded from the calculation as they would not have reached their full linguistic potential yet. Equation (76) demonstrates the calculation of the numerator C. Here P_i is the number of adults in group i, where each group i is unique with respect to the languages spoken. For instance, one group may speak only Afrikaans, English, Setswana and Sepedi. There are m groups altogether. The first term on the right-hand side of equation (76) gives the total number of possible conversations within group i. By definition, everyone in group i should understand everyone else in group i. The last term in equation (76) is the number of possible conversations between the individuals of group i and all other individuals in all other groups.

Clearly, this would depend on the degree to which language ability overlapped. The variable *d* is a 0-1 dummy variable where 1 means that because group *i*1 and group *i*2 share at least one language, they are able to communicate. If *d* is equal to 1 then the number of possible conversations is the product of the values P_i from the two groups. It is necessary to add the number of possible conversations across all possible two-way combinations of the *m* groups. The determination of *d* in each instance involves looking at the availability of languages in both groups *i*1 and *i*2. This lends itself to the running of programming code iteratively. Using the data and method discussed above, the value of *e* from equation (74) for South African adults was found to be 0.34. It was also possible, through considering which languages are spoken as home languages, to calculate permutations of *e*. For instance, if one only considers home languages, the value of *e* becomes 0.86. If one relaxes this restriction somewhat and allows for situations where at least one party in the conversation had to be using his home language as the language of communication, then the value of *e* is 0.60.

For the purposes of the growth simulator, what is important is whether enthnolinguistic fractionalisation can be expected to change over time. One way of attempting to answer this question is to examine whether the linguistic fractionalisation index for South Africa is different for older and younger adults. The next graph illustrates index values for five different age groups where, for instance, the age 20 value is based on observations with the age value in the range 15 to 24. There appears to be a clear improvement over time of around 0.05 every ten years. The flat trend between age 20 and 30 is likely to be a reflection of the fact that younger adults, specifically those aged 15 to 24, would not have realised their full linguistic potential in terms of languages spoken.

Figure 29: Linguistic fractionalisation in South Africa by age of adults



The fact that the average young South African adult was not able to communicate with around 30% of fellow South Africans due to language barriers in 2004 (this is the meaning of the index value) is a fragmentation problem that is clearly amenable to policy interventions, in particular education policy interventions. As suggested by Table 20, the specific solution seems to be a stronger emphasis on the most widely spoken language, English, with the aim of ensuring that all adults became fluent in this language. In 2004, despite being by far the most widely understood language, English was still not understood by 30% of adults. In a situation where all adults understood English, the linguistic fractionalisation index value discussed here would drop to zero. The trend, suggested above, of a reduction in linguistic fractionalisation equal to around 15% of the 2004 level every ten years, can be roughly transposed to the scale of the index used by SDM, or Alesina *et al* (2003), remembering that those two sources use scales which are almost identical. This would result in a drop every decade of around 0.12 points within a scale where South Africa's original fractionalisation was 0.88. This possible magnitude of change was considered possible for the purposes of the growth simulator. Of course this proposed magnitude is only slightly better than an educated guess, yet it seemed justifiable to come up with this magnitude given, firstly, that some empirical backing has been provided, albeit just for one country and, secondly, given that ethnolinguistic fractionalisation has been found to be a significant predictor of growth whilst at the same time it is a variable that is amenable to the influence of education interventions.

Government consumption share. Here the three equations (69) to (71) were used. The Penn data available went back to 1950 for many countries.

Rule of law. The three equations were applied to the annual values available for the 1996 to 2010 period for a large number of countries.

Total fertility rate. The three equations were applied to values for every five years going back to 1950.

Once plausible best trends for all eight explanatory variables had been established for the growth simulation tool, it became necessary to establish each variable's additive contribution to growth, using as the point of departure equation (73) above. The variable requiring the most complex calculation was educational quality given the lag between changes in the schooling system and changes in the human capital composition of the labour force. The following equation describes the approach taken to expand on equation (73).

$$\Delta g_Y = B \times \left(\frac{T_{t-1} \times 40 - Z_{t-1} + A_{t-1}}{40} - T_{t-1} \right) \qquad (77)$$

T represents the average quality of human capital in the labour force. The previous year's *T* is adjusted using *Z*, being average *T* amongst adults who had been in the labour market for 40 years in the previous year and would no longer be present in the current year and using *A*, being average *T* amongst young adults who entered the labour market between the previous and the current years. The result is *T* for the current year *t*.

The slope coefficient *B* for each of the explanatory variables was needed to obtain forecast estimates. Moreover, confidence intervals for each slope coefficient *B* at a 5% level of significance were needed in order to estimate the lower and upper bounds in the forecast. These statistics were obtained by using the results of existing growth regressions. Running a new growth regression using just the set of eight explanatory variables used in the growth simulation tool would have been ideal, though, as argued below, is not necessary. The new growth regression would have been difficult to run for a number of reasons to do with the availability of data for the start of the growth period, say around 1960. The problem of a lack of early data on educational quality has already been discussed (see section 4.4), and a new model would inevitably have repeated this problem found in Hanushek and Woessman (2009). Moreover, early data on the rule of law seems not to be readily available. Hanushek and Woessman (2009: 9) do include the related variable 'property rights', but it is not clear what data source they used or how early the values used in the growth regression were. Not running a new growth regression was not considered a great loss for the process of arriving at scenarios that might inform policymakers, given that those scenarios are aimed at providing greater empirical backing for policy priorities that are already more or less embraced by policymakers, as opposed to overturning existing intuition or advocating a radical re-prioritisation for governments. Had the latter, more radical intention been the case, a more rigorous approach that might have included a new growth regression would have been justified.

The slope coefficient *B* used for educational quality in the simulation tool was 0.007, which is taken from Table 16 (the Table 16 coefficient had to be divided by 100 given the metric of the educational quality variable in that model). Thus in the tool every increase of one point on the PISA quality scale results in a 0.007 addition to economic growth, so an increase of 100 PISA points, or one PISA standard deviation, might raise annual growth from 3.0 to 3.7 per cent. The size of the conditional correlation between educational quality and growth in the tool is thus around a half of what it is in the models of OECD (2010a). For rule of law and total fertility rate, again coefficients from Table 16 were used. For these two variables, the data sources in the tool are the same as those of the Table 16 regression, though in the case of the total fertility rate 1960 values were used for Table 16 whilst 2010 values were used in the tool. The assumption is thus made that the conditional correlation between the fertility rate and growth has not changed over time. In the case of rule of law, values are from 2000 in both Table 16 and the tool, meaning that criticisms similar to those applicable to the educational quality values, which are also recent, would apply.

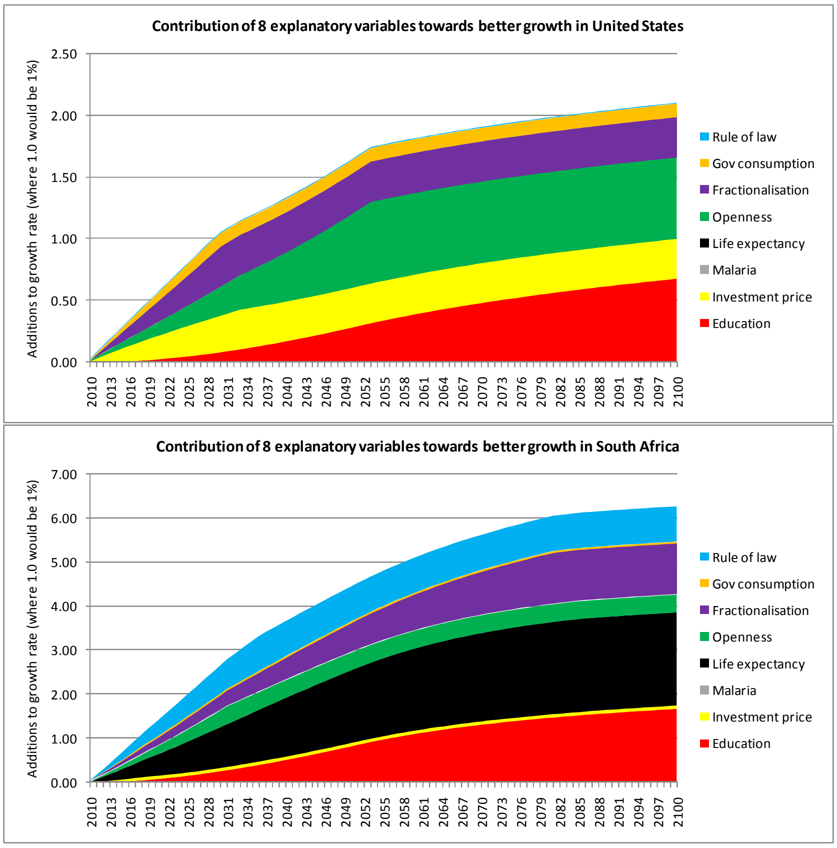
Because the simulation tool, unlike the projections in OECD (2010a), includes a large range of developing countries, it was especially important to allow for a large range of baseline growth rates and diminishing growth as countries developed. In the case of OECD (2010a) the approach was to apply a baseline 'education-as-usual' growth rate of 1.5% across all, mainly developed, countries. For the tool, each country's actual average growth rate for the period 2000 to 2009 was used for the baseline. Moreover, it was assumed that the growth rate would decline by one percentage point for every increase of 1.0 in the log of income between the years t_{-21} and t_{-20}. This means that a coefficient of -1.0 was used where the model in Table 16 yielded a value of -1.62 and SDM provided a coefficient of 0.9 on the log of previous income (see Table 10[53]). In other words, a point between the two empirical results was chosen. This choice seemed to result in plausible forecasts in the tool, though for certain countries anomalies arose, as discussed below.

Projections produced by the tool were broadly similar to those of OECD (2010a), with one exception: discounted present income values associated with the education intervention. This matter is also discussed below.

[53] The value -9.E-03 under 'Original SDM' must be multiplied by 100, producing 0.9, to take into account the fact that the model informing Table 10 treats the growth rate as a fraction.

Figure 30 below includes two graphs taken from two different runs of the tool, for the two countries United States and South Africa. In the case of the United States, improvements in educational quality result in, for instance, a growth rate in 2100 that is 0.7 percentage points greater than it would be without the improvement. This is relatively close to the corresponding 1.01 long-run addition to the growth rate for the United States seen in OECD (2010a: 25). In the case of South Africa, the addition to the growth rate by 2100 is considerably larger, at 1.7 percentage points. This is to be expected, given the larger scope for educational quality improvement in South Africa. One problem with the tool is that absolute per capita income figures are often implausible. For instance, the growth figures of Figure 30 produce a per capita income level of around USD 175,000 in the United States in 2100, and just over one million USD in the case of South Africa. Clearly this has little meaning for policymakers. What does seem meaningful, however, is the illustration of the kinds of interventions that stand out as particularly valuable in South Africa. These interventions are, apart from improving educational quality, improving life expectancy, which in South Africa would largely involve tackling HIV/AIDS, and dealing with linguistic fractionalisation. Tackling HIV/AIDS is clearly part of the national development plan of South Africa (South Africa: National Planning Commission, 2012), but the matter of a language policy that promotes social cohesion is not. One value of the picture provided by the South Africa graph below would thus be to bring to the fore a policy issue whose importance has been under-estimated.

Figure 30: Contributing factors behind higher future growth



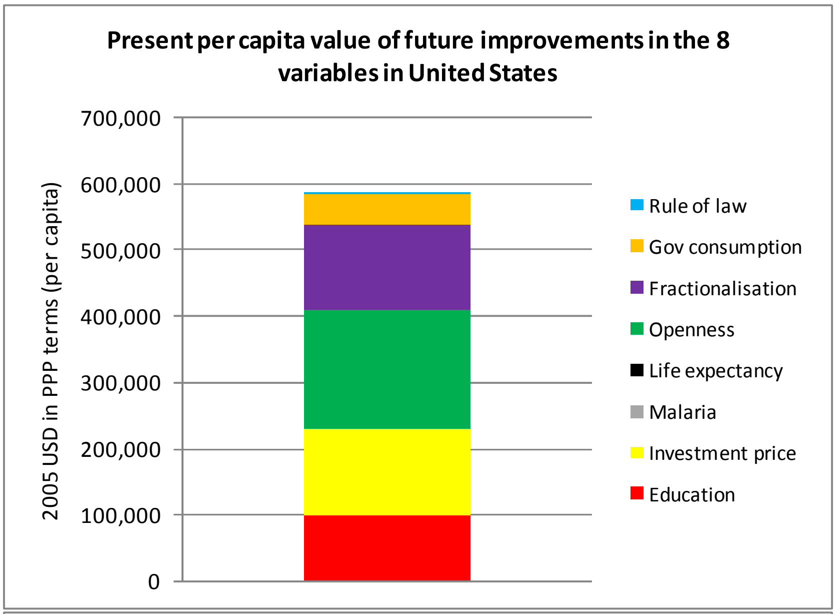
To calculate the discounted present value of income gains, the tool used a 3% discount rate and thus the same rate as the OECD (2010a). Results for the United

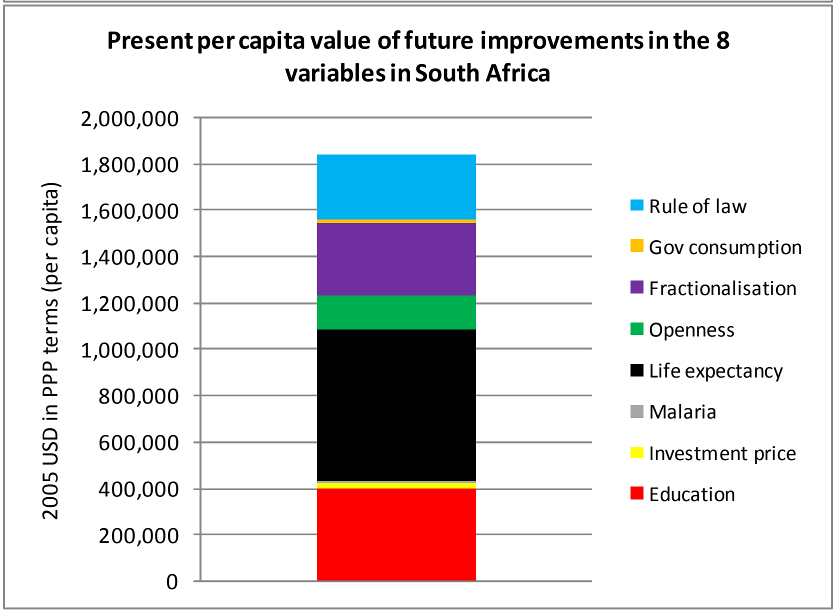
States and South Africa are shown in Figure 31 below. Again, a comparison across countries of the absolute values does not seem meaningful but what appears important, or at least suggestive, is the relative size of each segment within one country.

There was a notable discrepancy between the tool and OECD (2010a). Present discounted values produced by the tool were much lower than those reported in OECD (2010a). It appears as if in all three scenarios of OECD (2010a) the present discounted values are inconsistent with the model parameters presented in the report. The values are two to three times as high as they ought to be. For example, in the first scenario the present value of future income for the United States is 267% of the baseline GDP[54] (the OECD report focuses on total GDP, not GDP per capita). In Figure 31, the education segment for the United States equals 247% of initial GDP per capita, but here a far more ambitious educational improvement trajectory is envisaged, compared to what is found in the first scenario of OECD (2010a). A replication of the calculation in OECD (2010a) revealed that the present value of future gains should not be greater than 120% of the baseline GDP. The OECD (2010a) present values would be correct if a discount rate of around 1.5% and not 3.0% were used. The authors do say that they tried various discount rates, and make specific reference to Stern's 1.4%. It is possible that the present values in OECD (2010a) are based on a discount rate of around 1.4%, and not the stated 3.0% rate.

[54] Table 2 in OECD (2010b: 25) provides the information needed to gauge the baseline GDP value used.

Figure 31: Present value of future changes





5.6 Educational improvement and teacher pay

The foregoing discussion has concentrated largely on how the relationship between educational quality and economic growth can be conveyed to policymakers in a manner that brings to the fore the welfare costs of not taking full advantage of current opportunities to improve learning outcomes. But the space that governments have to improve education outcomes is often severely constrained by the demands of teacher unions. One of several constraints brought about by teacher unions is that they divert funding away from non-teacher inputs in the schooling process and towards teacher salaries, in a manner that results in allocative inefficiencies, and thus sub-optimal outcomes relative to the available budget. This problem is analysed empirically by Pritchet and Filmer (1997), who take coefficients from various developing country production functions, combine these with the unit costs of the production inputs, and examine the marginal gain in test scores associated with various investments. They find that the marginal gains from investing in teacher salaries are particularly poor, which points to a skewed and inefficient basket of inputs. In a situation of allocative efficiency, the marginal returns to the various investments should be more or less equal.

One way of extending the foregoing work with a view to making salary negotiations between the government and teacher unions a more informed and welfare-oriented process, would be to model the relationship between educational improvement and teacher pay. If teachers could see in fairly explicit terms the relationship between what they can influence, namely educational outcomes, and their pay they might focus more on outcomes, which would benefit not just teachers but society as a whole. This is the idea that informs the following series of graphs. As will be seen, the fact that there are particularly long lags between educational change and economic effects, and given that teacher union leaders tend to focus on fairly immediate gains, the effect that differently presented information can have on political processes is limited. Yet the analysis that follows, which appears not to be available in the existing literature, seems to be a valuable ingredient in any teacher wage negotiation process.

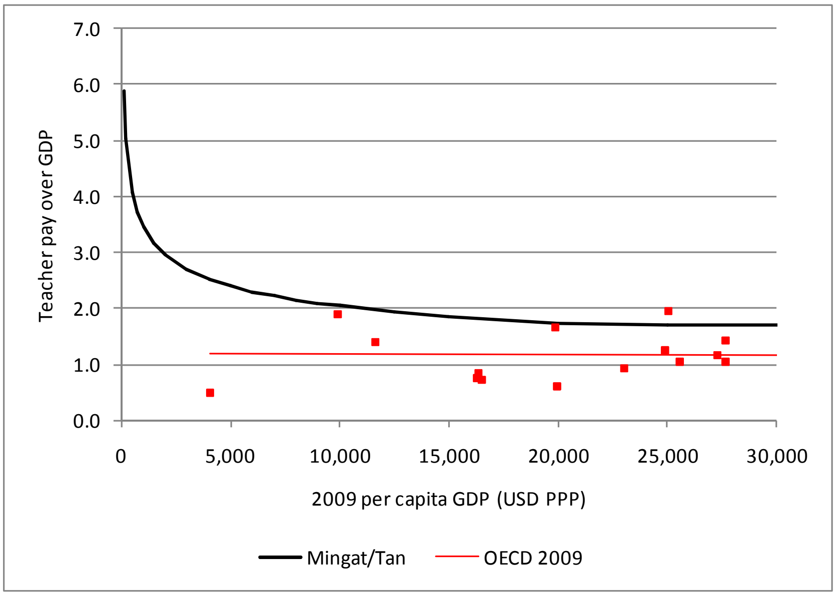
One point that must be emphasised is that it is not reasonable to assume that teacher pay will increase as fast as GDP per capita. Before the possible impact of educational improvement on teacher pay is examined, the relationship between teacher pay and GDP per capita must be explained. Mingat and Tan (2003), in an article titled ‘On the mechanics of progress in primary education’, provide valuable guidance on the latter relationship. For countries such as South Africa, where secondary school teacher salaries are very similar to primary school ones, this guidance is applicable to the schooling system as a whole. In a nutshell, the ratio of teacher pay to GDP per capita for a country declines over time, as the latter variable improves. It does this because as society develops and more professional positions outside the teaching profession become available, teachers become less of an elite in society and their relative, but not absolute, income declines.

Figure 32 below includes a reproduction of the Mingat and Tan (2003: 457, 466) curve relating teacher pay over GDP per capita to GDP per capita[55]. Equation (78) below describes the Mingat/Tan curve in the graph for ratios of teacher pay over GDP per capita, or *r*, down to 1.7. The variable *I* is income per capita. Mingat and Tan (2003) seem to suggest that in the case of developed countries, the ratio does not decline beyond around 1.7.

$$r = 17 I^{-0.23} \tag{78}$$

[55] The fact that GDP is used here whilst GNP is used in the original has almost no effect. The two variables produce a correlation coefficient of 0.9999 at the country level, using 2009 values from World Bank DataBank. The horizontal axis was recalibrated so that Figure 32 would reflect 2009 USD PPP values, based on Heston, Summers and Aten (2011).

Figure 32: Relationship between relative teacher pay and GDP per capita



Source: Mingat and Tan (2003: 461), with adjustment to 2009 price levels, and http://www.oecd.org/edu/skills-beyond-school/48631286.pdf. Red points refer to countries in the OECD source.

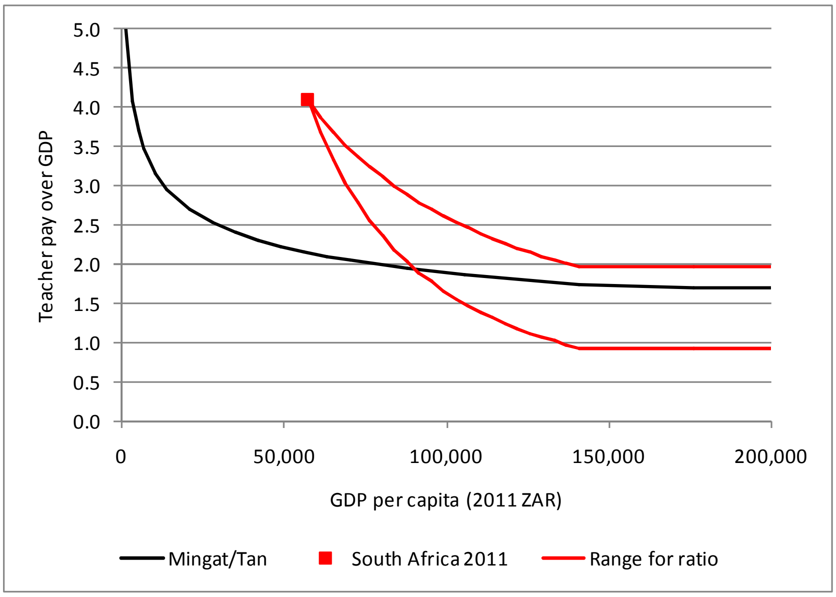
It would have been ideal to redo the Mingat and Tan analysis, but comparable average teacher pay values, in particular for developing countries, are not readily available. Mingat and Tan had to piece together information from a multitude of sources. What is relatively straightforward, however, is to examine teacher pay values for just OECD countries, as these are systematically published[56]. The average ratio at the developed country stage is in fact particularly important as one assumes that this is the ratio towards which developing countries such as South Africa should move. In Figure 32, the red curve and points reflect the situation in OECD countries plus one non-OECD country, Indonesia, the only non-OECD country for which the OECD analysts succeeded in finding data. Amongst OECD countries there is no negative or positive trend, which seems to confirm the existence of a steady state ratio in developed countries. What is noteworthy is that the OECD average for the ratio, at 1.16, is considerably below what is seen in Mingat and Tan (2003). The explanation for the difference appears not to be a downward trend over time in the ratio. According to the OECD source, the ratio remained constant, at least for the 2000 to 2009 period. It seems more likely that different definitions lie behind the difference. Mingat and Tan (2003: 458) use the 'average teacher salary' at the primary school level. The OECD source uses the average salary paid to teachers at the primary school level who have worked for 15 years and have just the minimum qualifications. It is likely that many teachers would have more than the minimum qualifications after being in service for 15 years.

A graph such as Figure 32 offers important guidance for policymakers and could help a government explain to teacher unions why teacher pay must in general increase at a lower rate than GDP per capita. But different developing countries, even ones with similar GDP per capita values, have vastly different ratios of teacher pay over GDP per capita. Overall, the range in the ratio for developing countries according to Mingat and Tan is 0.8 to 15.0. It would be useful to construct a graph that indicates the

[56] Information on OECD indicator D3 at http://www.oecd.org/edu/skills-beyond-school/48631286.pdf.

trajectory for a specific developing country from its current position to a future point at which it becomes a developed country. This is attempted in Figure 33 with respect to South Africa. It is assumed that when GDP per capita reaches 20,000 USD PPP (for 2009) a fixed band of possible ratios apply. That band in Figure 33 lies between 0.93 and 1.96, which are the 10^{th} and 90^{th} percentile values in the 2009 OECD data. The band thus encompasses the Mingat/Tan steady state ratio of 1.7.

Figure 33: Evolution of the teacher pay ratio in South Africa I



In the above graph, 2009 USD PPP are converted to 2011 ZAR (South African rand) values. The average pay for a teacher who has been in service for 15 years and has just the basic qualification was ZAR 235,510 in 2011[57]. GDP per capita in 2011 was ZAR 57,683[58]. This produces a ratio of 4.08, which lies well above the Mingat/Tan curve. If one assumes that a power function such as that expressed in equation (78) best describes the trajectory towards a lower ratio, then one could use a series of equations such as the following to describe the trajectory for a specific country with a specific point of departure. Here the subscripts *1* and *2* refers to the start and end points, where the end point is the entry point into developed country status at a GDP per capita of 20,000 USD PPP (in 2009) or ZAR 140,730 (in 2011). To illustrate, for the calculation of the lower bound, r_1 is 4.08, *r2* is 0.93, I_1 is ZAR 57,683 and I_2 is ZAR 140,730.

$$r_I = xI^y \tag{79}$$

$$y = \frac{\ln(r_2/r_1)}{\ln(I_2/I_1)} \tag{80}$$

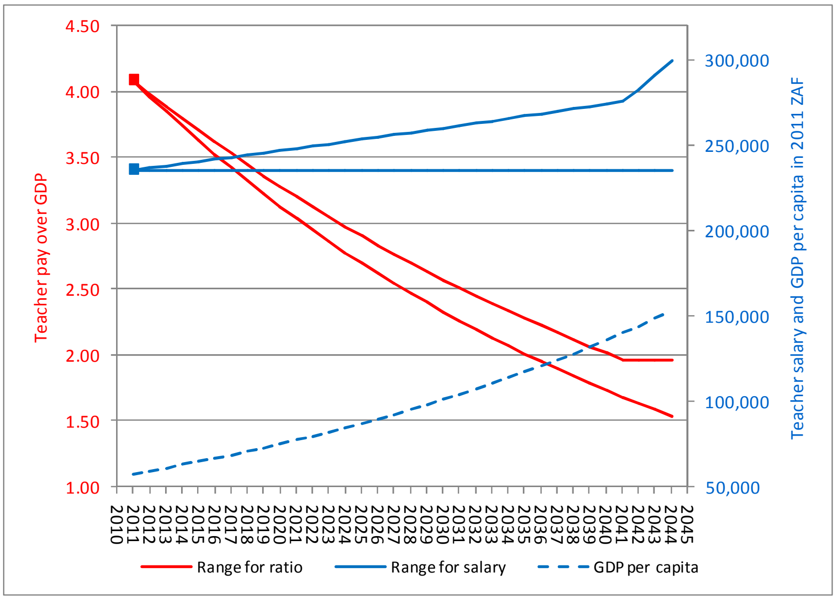
[57] Own analysis of October 2011 Persal payroll data accessed through the Department of Basic Education, Pretoria.

[58] *2013 Budget Review* available at http://www.treasury.gov.za for GDP and 'Mid-year population estimates' for 2011 available at http://www.statssa.gov.za.

$$x = \frac{r_1}{I_1^y} \tag{81}$$

For many, what would be more meaningful than the previous graph, is a picture of the evolution of actual teacher pay over time, given certain assumptions. This picture is provided in Figure 34 below. A growth rate for income per capita of 3.0% is assumed here, which is the average for the 2000 to 2009 period in South Africa. What becomes clear when one deals with actual teacher pay, is that the lower bound for the ratio illustrated in the previous graph can result in a decline in teacher pay. It is assumed that this is not politically possible and hence the lower bound is raised to ensure that teacher pay never declines below its initial level. This explains the narrowness of the range between the red curves in Figure 34 below. With respect to actual teacher pay, what is clear is that with a growth rate of 3.0%, only relatively modest real increases of around 0.5% per annum in teacher pay are possible for many years. It is only several decades into the future that larger annual increases for teachers become possible. This is a sobering picture, and would be counter-intuitive for those who adopt the assumption, probably a fairly common one, that teacher pay can rise at the speed of GDP per capita.

Figure 34: Evolution of teacher pay in South Africa II

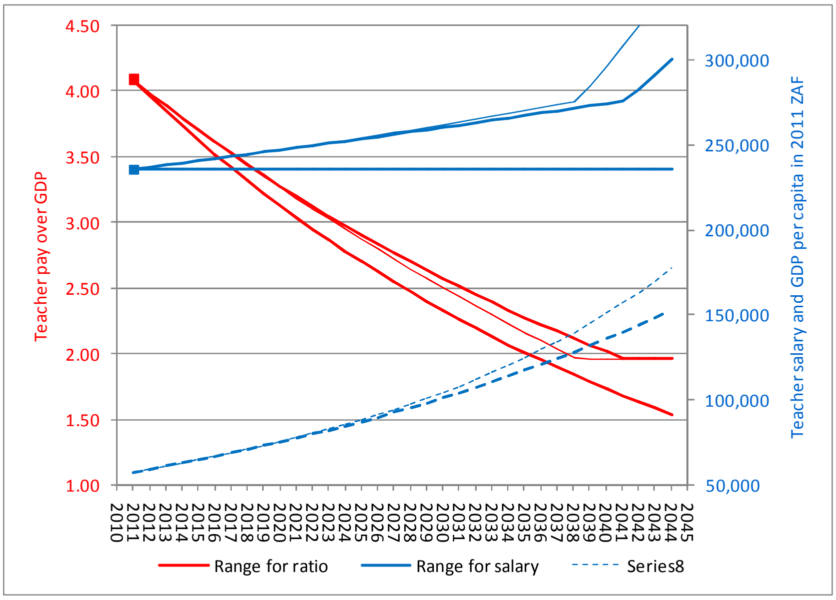


Importantly, an increase of 0.5% a year would not necessarily mean adjustments to the teacher pay scales. As pointed out by Gustafsson and Patel (2009) with reference to South Africa's teacher pay, what teachers currently earn is often different to what the current pay scales say teachers will earn, say, fifteen years into the future. This is because improvements in future earnings prospects built into the pay scales are generally not implemented retroactively. This dynamic exerts a potentially large impact on pay projections. The analysis by Gustafsson and Patel (2009: 19) suggests that simply by letting existing pay scales take effect, the pay of 45 year old teachers would increase by around 30% in real terms over a couple of decades.

A further complexity, analysed by Gustafsson (2012c) using South Africa data, is that if for historical reasons the age distribution amongst teachers is skewed either towards the high end or the low end, average teacher pay, across all teachers, may fall or rise purely on the basis of demographic trends. Specifically, in South Africa an ageing teacher force means that a large inflow of young teachers in the near future is likely, which would exert a negative impact on average teacher pay given that younger teachers are paid less as they have not accumulated the income benefits associated with years of experience. But if this dynamic is considered in combination with the pay scale effect of Gustafsson and Patel (2009), the net negative impact is small. The projection tool coming with Gustafsson (2012c) suggests that if one assumes recognition of experience increments in the future that are more or less in line with the current practice, there would be a period of decline in the average pay, between 2018 and 2035, but the overall decline would only be around 0.6% in real terms. These complexities serve as a reminder that a picture such as that in Figure 34 can be made more realistic through the incorporation of the actual teacher scales and demographic factors. A further complication that should ideally be taken explicitly into account is the fact that thresholds between different levels of development, such as between developing and developed country status, change in complex ways that are not simply a function of inflation. This would influence the point at which the developed country ratio of teacher pay over GDP per capita would take effect. How the relationship between income per capita, in constant price terms, and a country's development status changes over time is discussed in Part II of the dissertation.

As much of section 5 has shown, GDP per capita is not an entirely exogenous variable as far as the education planner is concerned. Education planners and teachers, through their actions, can exert a positive impact on GDP. The following graph reproduces the curves of the previous graph, but adds new and higher upper bounds for GDP per capita and teacher pay that might result from improvements in educational outcomes. The upper bound for the ratio of teacher pay over GDP per capita would drop, however, given the faster move towards the developed country threshold. Using the education quality coefficient of 0.007 described in section 5.5, meaning 100 PISA points would raise annual growth from, say, 3.0% to 3.7%, results in barely noticeable changes to income and hence teacher pay. What seems clear is that education planners cannot honestly say to teachers 'Improve the performance of your students and government will be in a position to pay salaries that are substantially higher in the near future'. Even if the education quality coefficient is doubled and thus put on a par with the coefficient used by Hanushek and Woessman in OECD (2010a), the improvements in teacher pay appear fairly unimpressive. In fact, Figure 35 below is based on an inflated coefficient of 0.014. In this scenario, it is only after 23 years that better educational quality will produce a GDP per capita level that will allow teacher pay to be at least 1.0% higher than what it was in the scenario painted in the previous graph, Figure 34.

Figure 35: Future teacher pay in South Africa with quality improvements



The calculations used for the above graph are explained in the equations below. Largely, they represent an unpacking of the earlier equation (77), but here the fact that growth diminishes as income rises is reflected explicitly. Growth *g* in year *Y* is the previous year's growth plus an education effect *e* and a convergence effect *c*. The education effect *e* is the educational quality coefficient *B* multiplied by the change in the quality of the labour force *T*. For the above graph, *B* was 0.014. The educational quality of the labour force in the current year is a function of the previous year's overall value, the value of the oldest age cohort of workers, *Z*, and the value *A* for the students from the previous year that became the current year's youngest workers. *Z* for the above graph, and for any scenario involving fewer than 41 future years, will be the average level of educational quality in the schooling system in the baseline year. *A* will rise annually, assuming optimal education interventions, according to Figure 23 above, whose values are reproduced in Table 22 in the appendices. The convergence effect *c* is the coefficient *E*, whose value is assumed to be -1.00 as discussed in section 5.5, multiplied by a value which is the natural logarithm of per capita income 20 years ago minus the same value from 21 years ago

$$g_Y = g_{Y-1} + e_Y + c_Y \tag{82}$$

$$e_Y = B(T_Y - T_{Y-1}) \tag{83}$$

$$T_Y = \frac{T_{Y-1} \times 40 - Z + A_{Y-1}}{40} \tag{84}$$

$$c_Y = E(\ln(I_{Y-20}) - \ln(I_{Y-21})) \tag{85}$$

5.7 Conclusion

One aim of section 5 was to explore how economic models can be meaningfully presented to policymakers. Findings in this regard have been implicit rather than explicit in the foregoing discussions. This conclusion attempts to bring these findings to the fore.

What seems clear is that making the academic economics texts more accessible requires an investment in what one could refer to as bridging texts and tools. As the literature on the relationship between academia and policymakers makes clear, there is a shortage of materials that can be used for knowledge brokerage between the two spheres.

So what makes for good bridging materials? Firstly, they need to draw from a variety of sources in the academic literature. Put differently, they need to function as meta-analyses. The growth simulation tool described above drew from two of the sources discussed in section 4, as well as a few other texts.

Secondly, they need to focus rather strongly on identifying key patterns within complex systems. Policymakers are inevitably confronted, on a continuous basis, by a large amount of new information and appeals from lobbyists. Academics are privileged insofar as they have, or should have, considerable space for questioning and contemplation. Research ought to assist in cutting through the mass of information, in ensuring that one does not miss the wood for the trees. An important part of the ‘wood’, is the likely long-run impacts of present actions and policies. Inbar (1996: 40), in an important education planning manual of the International Institute for Educational Planning (IIEP), says the following about the importance of identifying the ‘wood’:

> Vision for educational innovation should yield a description that is useful to the viewer and not cluttered by irrelevant information. Vision is built around certain familiar symbols and is itself a creator of new symbols.

Importantly, there are two levels to the problem. First, analysts need to have a firm grasp of the ‘wood’ or the key patterns that must be understood if effective policies are to be formulated. Secondly, they need to be able to convey this knowledge in ways that are intelligible to the policymaker. The economics discipline is well placed to offer guidance at both levels for a number of reasons. The economics literature often pays very close attention to how one measures and quantifies things. It is thus rather rigorous in the way it concludes empirically what current and past realities are. Misunderstandings of these realities are frequently a problem in education policymaking. The economics literature also pays considerable attention to identifying, measuring and understanding correlations, including counter-intuitive ones, between different variables. It teases out from an enormously complex and often data-rich system what factors seem to matter for the key variables of interest, such as learning outcomes. It does this within a conceptual framework with a long tradition and whose key concepts, efficiency, equity, welfare, costs and benefits, are of great relevance to policymakers.

Yet there is a risk that too much is promised, that caveats are not made sufficiently clear, in the quest for elegance and simplicity. Inbar (1996: 91) makes the following warnings:

> Comprehensive knowledge, organized into a coherent framework, gives the appearance of better control, or reduced uncertainty, and a decrease of risks. The development of over-sophisticated, abstract models and plans of innovation arises from this illusion. Modelling is often undertaken to persuade politicians or administrators who demand certainty before allowing planned change and who too often search for perfection rather than improvement. ... [models] are imbued with almost magical quality of truth and objectivity. Consequently, there is pressure to advocate rationality, which may contribute little or nothing to the change process, rather than into the development of human contacts, involvement and commitment which are true underpinnings of innovation.

A third priority is thus to avoid merely playing to the 'baser instincts' of policymakers and, more pertinently, politicians, who, perhaps as a result of exhaustion induced by information overload, may want to settle for an analysis that is overly simplistic. In section 5, acknowledgement that different analysts have tackled similar policy problems differently and the use of confidence intervals in the modelling of scenarios are both ways of counteracting the illusion that the future can be perfectly modelled.

Fourthly, cross-country comparisons should be employed, but in a manner that acknowledges the serious pitfalls in this area. These comparisons, which lie at the heart of the key models discussed in this dissertation, are useful to policymakers, who are in fact often asked to explain the effectiveness of the policies of their own country in terms of what is found elsewhere. Advice to policymakers should include reference to common international 'league tables', such as those of the World Economic Forum and the World Bank. However, this advice needs to acknowledge the serious measurement problems in the international indicators. Some of these problems were discussed in section 5.4. In fact, it is in this regard that economists should display some humility, for it is often economists who are responsible for the league tables in question. It is important to explain to the policymaker the distinction between more academic texts produced by economists, such as peer-reviewed articles, and reports, also produced by economists, where, it could be argued, rigour has been compromised so that an appearance of comprehensive knowledge could be maintained.

A fifth point is that when one tries to bridge the existing academic literature with the needs of policymakers, one tends to find key gaps in the literature, because that literature was often not produced with a clear understanding of what policymakers might require. These gaps need to be filled where possible. In section 5, one such gap was insufficient analysis into what historical trends tell us about feasible future improvement trajectories in the area of average test scores. Another gap was the missing answer to the question as to whether linguistic fractionalisation in a society such as South Africa's changes over time and, if so, at what speed. Both these gaps relate to target-setting, something that policymakers are increasingly concerned with as governance through indicators becomes more common.

The sixth point is that the bridging texts and tools need to be designed in a manner that is sensitive to who the potential users are and, in the case of policymakers in government, typical within-government dynamics and dynamics between the government and important stakeholder groups, for instance teacher unions. The growth simulator tool discussed above expanded on the work of OECD (2010a) through the incorporation of various development indicators because at least some education policymakers rightly see their role as lying at the heart of the long range development of the country. The attention paid to the ethnolinguistic fractionalisation indicator was largely justified through reference to its sensitivity to the education process. Policy on the overall language situation in a country is typically not well accommodated within traditional government structures, despite its apparent importance for development, yet if there is a home for this policy area, it is arguably within the education ministry. The importance for the budgeting process of teacher pay, and the political weight of teacher unions, prompted the extension of the analysis to encompass long range teacher pay trends in section 5.6. Admittedly, what is presented above in relation to teacher pay is far from a comprehensive model of the long-range relationship between teacher pay and economic growth. As was explained above, such a model would need to take into account more operational realities relating to factors such as the teacher pay scales and the age distribution of teachers. The emphasis was on putting forward some broad analytical checks and balances that would need to be considered when the evolution of teacher pay is related to growth.

6 CONCLUSIONS FOR PART I

Part I of the dissertation contains a mix of literature reviews, often enhanced by re-estimations or re-presentations of existing models, as well as what can be considered new findings based on new data analysis. What have the key re-estimations, re-presentations and findings been, and what is their importance in terms of the research questions put forward in the introduction in section 1?

Section 2, which outlined around fifty years of thinking on economic growth, has in a sense been put together as a primer for education planners on this important area of economics. In attempting to simplify a rich and complex body of literature, the various traditions were condensed into standard graphs to facilitate comparison and a clearer sense of the evolution of ideas over time. Importantly, the theoretical literature in question is not entirely abstract and removed from practical policymaking concerns in the education sector. It deals with key concepts such as the trade-off between imitation and innovation and the role of time in the education process, concepts that good education policymakers would be grappling with.

In contrast to the non-empirical theory of section 2, section 3 dealt with concrete questions around how to measure both educational outcomes and what we want these outcomes to enhance, namely human welfare. Again, the focus was strongly on what education policymakers need to know. It was assumed that these policymakers do need to understand the complexities around measuring welfare, but also that a variety of indicators of welfare exist which, if interpreted with care, are useful. If education does indeed lie at the heart of advancing welfare, as much of the theory suggests, then surely those who design the education system must have a good grasp of the key accomplishments relating to the measurement of welfare. The economic literature on the likely impacts of climate change on welfare was introduced, as out of this literature is emerging some of the most critical policy advice that policymakers have ever had to deal with. Education policymakers are only beginning to think about how the education system impacts on both climate change mitigation and adaptation. But beyond these practical concerns, school curriculum specialists will have to pay increasing attention to how to package ethical concerns relating to the environment within the school curriculum.

Section 3.3 includes one of the innovations put forward by Part I of the dissertation. Here a method was proposed for constructing a set of country-level education quality measures that is wider than the sets of values existing previously. Specifically, the set covering 77 countries and put forward by Hanushek and Woessman (2009) was expanded on, using different methods and some new data, in order to come up with a set covering 113 countries. The new set has a greater presence of developing countries. As explained in section 3.3, the new method comes with some costs relating to measurement rigour, but it was argued that the new set of values is useful, both for cross-country modelling, and as an expanded ‘league table’ that can inform policy debates.

The discussion on theory, in section 2, and measurement, in section 3, can be considered necessary preambles to the examination, in section 4, of three prominent cross-country growth models (or more precisely, three sets of models). A substantial amount of the methodological focus went towards the partially Bayesian methods of Sala-i-Martin, Doppelhofer and Miller (2004), referred to as the BACE approach.

These methods seem valuable for data where there are relatively few observations and many potential explanatory variables. Whilst BACE does not establish cause and effect, it does make the discovery of conditional correlations a more scientific process. Cross-country growth regression data is clearly one candidate for BACE, but it could be applied to other areas of analysis that are of importance to education policymakers, such as education production functions. The key policy discussions in section 4 relate to the landmark analysis of Hanushek and Woessman (especially 2009). The work of Hanushek and Woessman is perhaps better known amongst education policymakers than any other recent work in economics, partly due to what Hanushek and Woessman say and partly due to how they say it. What they say has attracted the attention of policymakers because of the very clear correlations between country growth and educational quality that have been brought to the fore. But their work is also important insofar as they pay close attention to the question of cause and effect and use statistical techniques to confirm that it is better education that largely brings about higher income. Causality does not run in the opposite direction. But section 4 also cautions against a too uncritical treatment of Hanushek and Woessman's work. Legitimate concerns have been raised about how much certainty the available data allow. The conclusion was drawn that whilst tentatively the findings of Hanushek and Woessman can be considered valid, we shall have to wait for another couple of decades of educational quality data to be collected before we can be certain about what the magnitude of the impact of educational quality on growth has been.

Much of what is new in Part I of the dissertation is found in section 5, where the centra concern is making the available economic analysis on education and growth 'talk better' to education policymakers. Section 5 is accompanied by an Excel tool that simulates the impact of educational quality, but also other development indicators, on growth and produces a set of graphs applicable to the country selected in the tool. The tool on its own is of limited value, but together with the narrative of section 5, it is meant to stimulate and inform the policy discourse around education's role in country development. A further aim is to assist cross-country comparisons, something that policymakers generally consider important, but is often elusive because the texts and tools required for this are limited and, as argued in section 5, occasionally misleading. Three rather novel analyses are explained in section 5. The first involves the use of the available country-level data to provide a more rigorous analysis of what education policymakers can consider feasible education quality improvement targets. The second involves the use of an unusual South African dataset to improve on the existing definitions of linguistic fractionalisation in a country, a factor which is partly dependent on how education is implemented in the country and is considered important for growth. The analysis includes an exploration of how linguistic fractionalisation changes over time, something that seems not to have been covered in the existing literature. The third novelty involves bringing together the 'mechanics of progress' work of Mingat and Tan (2003), Hanushek and Woessman (2009) and some practical considerations of how teacher pay works in South Africa, in order to represent teacher pay in the national development context in a new manner which could inform the teacher wage negotiation process.

Appendix A

The following table contains adjusted programme-specific and across-programme country scores discussed in section 4.4.

The Excel tool developed to implement the adjustments is file *Cross-country.xls*, which accompanies this dissertation. For the solving of the nonlinear programming problem described in section 4.4 above, the Solver facility in Excel 2003, with default settings applicable, was used.

Table 21: Adjusted country averages of pupil performance

Country	*ISO code*	*PIRLS*	*PISA*	*SAC-MEQ*	*SER-CE*	*TIMSS Gr4*	*TIMSS Gr8*	*Overall*
Albania	ALB		373					373
Algeria	DZA					401	395	398
Argentina	ARG	401	391		392			395
Armenia	ARM					459	464	462
Australia	AUS		522			477	472	490
Austria	AUT	485	497			475		486
Azerbaijan	AZE		406					406
Bahrain	BHR						404	404
Belgium	BEL	477	513			502	497	497
Belize	BLZ	336						336
Bosnia and Herzegovina	BIH						442	442
Botswana	BWA			355			380	367
Brazil	BRA		381		392			387
Bulgaria	BGR	492	422				452	455
Canada	CAN	491	529			477	484	495
Chile	CHL		419		407		395	407
China (Shanghai)	CHN		578					578
Colombia	COL	403	387		388	387	390	391
Costa Rica	CRI		426		425			425
Croatia	HRV		470					470
Cuba	CUB				467			467
Cyprus	CYP	454				478	446	459
Czech Republic	CZE	484	495			464	475	479
Denmark	DNK	491	503			486		493
Dominican Republic	DOM				328			328
Ecuador	ECU				352			352
Egypt	EGY						403	403
El Salvador	SLV				370	373	363	369
England	ENG	491				493	476	487
Estonia	EST		507				493	500
Finland	FIN		543					543
France	FRA	474	501					488
Georgia	GEO	437	377			436	411	415
Germany	DEU	488	497			487		491
Ghana	GHA						331	331
Greece	GRC	475	463					469
Guatemala	GTM				353			353
Hong Kong	HKG	490	539			525	526	520
Hungary	HUN	491	487			484	488	487
Iceland	ISL	466	503					485
Indonesia	IDN	391	379				407	392
Iran	IRN	400				411	409	406
Ireland	IRL		506					506
Israel	ISR	465	448				458	457
Italy	ITA	491	473			475	460	475
Japan	JPN		523			511	520	518
Jordan	JOR		394				421	408
Kazakhstan	KAZ		398			501		449
Kenya	KEN			388				388

Country	*ISO code*	*PIRLS*	*PISA*	*SAC-MEQ*	*SER-CE*	*TIMSS Gr4*	*TIMSS Gr8*	*Overall*
Korea	KOR		542				535	539
Kuwait	KWT	361				365	373	366
Kyrgyzstan	KGZ		310					310
Latvia	LVA	488	478			494	478	484
Lebanon	LBN						432	432
Lesotho	LSO			288				288
Liechtenstein	LIE		516					516
Lithuania	LTU	486	475			491	475	482
Luxembourg	LUX	498	474					486
Macao	MAC		509					509
Macedonia	MKD	417	377				428	407
Malawi	MWI			260				260
Malaysia	MYS		409				466	437
Malta	MLT		453				464	458
Mauritius	MUS		414	417				415
Mexico	MEX		407		410			408
Moldova	MDA	455	393			475	445	442
Montenegro	MNE		401					401
Morocco	MAR	342				381	393	372
Mozambique	MOZ			332				332
Namibia	NAM			289				289
Netherlands	NLD	494	527			494	497	503
New Zealand	NZL	479	524			468	468	485
Nicaragua	NIC				361			361
Norway	NOR	457	497			450	448	463
Oman	OMN						385	385
Palestine	PSE						389	389
Panama	PAN		366		358			362
Paraguay	PRY				358			358
Peru	PER		339		373			356
Philippines	PHL					389	389	389
Poland	POL	472	492					482
Portugal	PRT		473					473
Qatar	QAT	354	343			353	341	348
Romania	ROU	458	419				450	443
Russia	RUS	491	462			494	479	481
Saudi Arabia	SAU						357	357
Scotland	SCO	477				468	467	471
Serbia	SRB		430				460	445
Seychelles	SYC			402				402
Singapore	SGP	488	544			529	539	525
Slovakia	SVK	475	483			470	478	476
Slovenia	SVN	466	496			467	470	475
South Africa	ZAF	318		322			311	317
Spain	ESP	467	480					473
Swaziland	SWZ			368				368
Sweden	SWE	497	506			474	469	486
Switzerland	CHE		514					514
Syria	SYR						388	388
Taiwan	TWN	483	521			513	534	513
Tanzania	TZA			385				385
Thailand	THA		422				432	427
Trinidad and Tobago	TTO	412	415					414
Tunisia	TUN		376			374	414	388
Turkey	TUR	422	441				426	430
Uganda	UGA			317				317
Ukraine	UKR					454	446	450
United Arab Emirates	ARE		426					426
United Kingdom	GBR		506					506
United States	USA	487	491			486	476	485
Uruguay	URY		425		427			426
Yemen	YEM					311		311
Zambia	ZMB			261				261
Zimbabwe	ZWE			345				345

Appendix B

What follows is the Mata code that was written to implement the Bayesian averaging of classical estimates (BACE) approach put forward by Sala-i-Martin, Doppelhofer and Miller (2004). The Guass code of those authors served as basis for developing the Mata code. The Mata code was developed to work in a Stata 9 environment (though it should also work in all subsequent versions of Stata).

```
* INTRODUCTION:

* This code implements the Bayesian averaging of classical estimates (BACE) methodology of Sala-i-Martin, Doppelhofer
* and Miller (2004). This is essentially a translation of these authors' code in GAUSS to Mata. This code was
* tested in Mata existing in Stata 9, but it was also found to work in Mata within Stata 12. The approach was to
* follow the variable naming of the original GAUSS code, unless renaming seemed to make things a lot clearer, and
* to have key comments from the original code plus new comments I thought were worth introducing.
* Importantly, this code produces only the three key statistics per variable from
* the original, namely the posterior inclusion probability (see macro 'postprob') below, the posterior mean conditional
* on inclusion ('coeff') and the posterior standard deviation conditional on inclusion ('sd'). This should be adequate
* for most analyses.

* RUNNING THE CODE:

* Select this entire file and then execute. Execution takes many hours, for instance 7.

* STARTING OFF IN STATA:

* If a dataset other than the original one is used, then it will be necessary to change the filename and the
* references to variable names in the following.
set more off
use "C:\BACEdata.dta", clear /* Should obviously change as needed. */
local xfirst "abslatit"
local xlast "ztropics"
drop obs code country

* The following ensures that there are no observations with any null values in the explanatory variables.
quietly foreach var of varlist `xfirst'-`xlast' {
  drop if `var'==.
}

* BEGIN USE OF MATA IN STATA:

clear mata
mata

/* LOAD DATASET WITHOUT MISSING VALUES AND POPULATE MATRICES CONTAINING THE DATA */

st_view(data_c=., ., .)
n = rows(data_c)
```

```
m = cols(data_c)
nvar = m - 1 /* Note that nvar is the number of explanatory variables, with y excluded. */
nobs = n
varnames = st_varname((2..m))
varnames = varnames'
y = data_c[., 1] /* The first variable must be the dependent variable. */
sdx = J(nvar, 1, .)
for (i=2; i<=nvar + 1; i++) {
  sdx[i-1,1] = sqrt(variance(data_c[.,i], 1))
}
sdy = sqrt(variance(data_c[.,1], 1))

/* SET MATRICES THAT KEEP THE SAME VALUE THROUGHOUT THE PROCESS */

/* Key prior assumption. */
pmsize = 7 /* Prior model size. */
/* Matrices relating to the weighting of ESS */
samplen = 100000 /* Frequency in loops of adjustments to the selection probabilities of variables. */
sprobb = 0.5 /* Partial adjustment coefficient for the sampling probabilities. 0 gives no adjustment, 1 gives immediate adjustment to the estimated posterior inclusion probability.
*/
lb = 0.1   /* Lower bound on the sampling inclusion probability. */
ub = 0.85   /* Upper bound on the sampling inclusion probability. */
/* Matrices needed for the monitoring of convergence */
checkn = 10000   /* Frequency in terms of number of loops for checking convergence. */
criterion = 10e-6   /* A threshold that is used to determine how stable the slope coefficients have become. */
numbertimes = 10   /* Another parameter used to assess convergence towards stable coefficient values. */
loopnum = 100000000 /* The number of loops that will occur if convergence is not reached. */
/* Matrices relating to the reporting of results */
semifinaln = 100000 /* Frequency in loops of reports on the values for the three key statistics. */
/* Matrix needed for the final calculation of posterior means of coefficient and standard deviation. */
tolerance = J(nvar, 1, 1)

/* SET MATRICES THAT CAN CHANGE DURING THE PROCESS */

/* The three key statitics found, per explanatory variable. */
postprob = J(nvar, 1, 0) /* Posterior inclusion probability. */
coeff = J(nvar, 1, 0) /* Posterior mean coefficient conditional on inclusion. */
sd = J(nvar, 1, .) /* Posterior standard deviation conditional on inclusion. */
/* The loop counters */
loop0 = 1 /* The counter that continues throughout with no resetting. */
```

```
loop1 = 1 /* The counter that checks when the next adjustment of selection probabilities should occur. Gets reset to zero. */
loop2 = 1 /* The counter that checks when the next interim report of statistics should be provided. Gets reset to zero. */
loop3 = 1 /* The counter that checks when the next convergence check occurs. Gets reset to zero. */
/* Matrix relating to the sampling inclusion probability. */
sprob = 0
/* Matrices relating to the value of ESS */
ess = 0 /* The ESS for just the current model. */
sess = 0 /* The cumulative sum of all ESS values. */
sumess = J(nvar, 1, 0) /* For each explanatory variable, the sum of the ESS values where the variable was included. */
/* Matrices needed for the monitoring of convergence. */
good = 0 /* Used in the assesssment of convergence. */
sflag = 0
cbeta = J(nvar, 1, 0)
maxchng = 1
avchng = 1
changer = 1
maxchngpp = 1
avchngpp =1
changepp= 1
postprobold = J(nvar, 1, 0)
sessold = 0
/* Matrices needed for the calculation of the posterior mean coefficients. */
coeffsum = J(nvar, 1, 0)
varw = J(nvar, 1, 0)

/* SET VALUES OF pprob (WHICH NEVER CHANGE) AND INITIAL VALUES OF iprob */

inclprob = pmsize / nvar
pprob = J(nvar + 1, 1, 0)
i = 0
for (i=0; i<=nvar; i++) {
  pprob[i + 1, 1] = inclprob ^ i * (1 - inclprob) ^ (nvar - i)
}
iprob = J(nvar, 1, 1) * inclprob /* This is the probability of being included in the model of each explanatory variable. */

/* SET SOME OTHER THINGS */
tolerance = tolerance * 10e-50 /* These values, rarely if ever used, cater for the possibility of an explanatory variable being virtually valueless in the analysis. */

/* START THE MAIN LOOP */
```

```
while (sflag==0) { /* Looping ends if the maximum loops is exceeded or convergence is obtained. */

/* SELECT THE VARIABLES FOR THIS LOOP */

  coinfliprnd = uniform(nvar, 1)
  coinflip = J(nvar, 1, 0)
  for (i = 1; i<=nvar; i++) if (coinfliprnd[i, 1]<iprob[i, 1]) coinflip[i, 1] = 1

  k = sum(coinflip)

  if (k>0) { /* We only enter the loop if at least one explanatory variable has been selected. */

/* CONSTRUCT THE SUB-DATASET */

    select = J(k, 1, .)
    x = J(nobs, k + 1, .)
    iin = 0
    for (i=1; i<=nvar; i++) {
      if (coinflip[i, 1]==1) {
                        iin = iin + 1
                        select[iin, 1] = i
                        x[., iin] = data_c[., i + 1]
                }
    }
    x[., k + 1] = J(nobs, 1, 1)

/* CALCULATE THE PROBABILITY THAT THIS MODEL WOULD BE SELECTED */

   if (coinflip[1, 1]==1) {
      sprob[1, 1] = iprob[1, 1]
    }
    else {
      sprob[1, 1] = 1 - iprob[1, 1]
    }
    for (i=2; i<=nvar; i++) {
      if (coinflip[i, 1]==1) {
        sprob = sprob * iprob[i, 1]
      }
                else {
        sprob = sprob * (1 - iprob[i, 1])
```

```
            }
    }

/* PERFORM THE REGRESSION ANALYSIS */

    invxx = invsym(x' x)
    beta = invxx * x' y
    resid = y - x * beta
    ess = (resid' resid) ^ (-nobs / 2) / nobs ^ ((k - pmsize)/ 2) /* This is the first BACE-specific adjustment of the residual sum of squares. */
    ess = ess * pprob[k + 1, 1] / sprob /* This is the second BACE-specific adjustment. */
    sess = sess + ess
    sumess = sumess + ess * coinflip
    coeffsum[select, 1] = coeffsum[select, 1] + beta[1::k, 1] * ess
            invxxdiag = J(k + 1, 1, .)
            for (i=1; i<=k + 1; i++) invxxdiag[i, 1] = invxx[i, i]
            varb = (resid' resid) * invxxdiag / (nobs - k - 1)
            varw[select, 1] = varw[select, 1] + (varb[1::k, 1] + beta[1::k, 1] :* beta[1::k, 1]) * ess

/* ADJUST THE SAMPLING PROBABILITY TO FAVOUR THE BETTER VARIABLES */

    if (loop1>=samplen) {
      loop1 = 0
      postprob = sumess / sess
      iprob = iprob + sprobb * (postprob - iprob)
      for (i=1; i<=nvar; i++) {
        if (iprob[i, 1]<lb) {
          iprob[i, 1] = lb
        }
        else {
          if (iprob[i, 1]>ub) {
            iprob[i, 1] = ub
          }
        }
      }

 /* DISPLAY THE NON-FINAL THREE STATISTICS PER EXPLANATORY VARIABLE AS THEY NOW STAND */

      if (loop2>=semifinaln) {
"Loops done:"
              loop0
```

```
            loop2 = 0
        postprob = sumess / sess
        for (i=1; i<=nvar; i++) {
                    coeff[i, 1] = coeffsum[i, 1] / max(sumess[i, 1] \ tolerance[i, 1])
                    sd[i, 1] = sqrt(varw[i, 1] / max(sumess[i, 1] \ tolerance[i, 1]) - coeff[i, 1] * coeff[i, 1])
                   }
"THE THREE STATISTICS:"
"1: Posterior inclusion probability"
"2: Posterior mean conditional on inclusion"
"3: Posterior standard deviation conditional on inclusion"
        semifinal = (postprob, coeff, sd)
        varnames
                   semifinal
      } /* The conditional on whether it is time to report on the interim statistics ends here. */
    } /* The conditional on whether to adjust the sampling probability ends here. */


/* CHECK FOR CONVERGENCE */

    if (loop3>=checkn) {
      loop3 = 0
      normcoeff = coeffsum :* sdx / sdy
      if (loop0>checkn) {
            chng = abs(normcoeff / sess - cbeta / sessold)
        maxchng = max(chng)
        avchng = mean(chng, 1)
        changer = .
        i = 1
        while (changer==.) {
          if (chng[i, 1] = maxchng) {
            changer = i
          }
          i = i + 1
        }
        postprob = sumess / sess
        chngpp = abs(postprob - postprobold)
        maxchngpp = max(chngpp)
        avchngpp = mean(chngpp, 1)
        changepp = .
        i = 1
```

```
        while (changepp==.) {
          if (chngpp[i, 1] = maxchngpp) {
            changepp = i
          }
          i = i + 1
        }
      }
      cbeta = normcoeff
      sessold = sess
      postprobold = postprob
      if (maxchng<criterion) {
        good = good + 1
      }
      else {
        good = 0
      }
      if (good>numbertimes-1 | loop0>loopnum) sflag = 1
    } /* The conditional on whether to run the convergence test ends here. */
  } /* The conditional on whether any explanatory variables have been selected ends here. */

  loop0 = loop0 + 1
  loop1 = loop1 + 1
  loop2 = loop2 + 1
  loop3 = loop3 + 1

} /* The main loop ends here. */

/* DISPLAY THE FINAL STATISTICS */

postprob = sumess / sess
for (i=1; i<=nvar; i++) {
  coeff[i, 1] = coeffsum[i, 1] / max(sumess[i, 1] \ tolerance[i, 1])
  sd[i, 1] = sqrt(varw[i, 1] / max(sumess[i, 1] \ tolerance[i, 1]) - coeff[i, 1] * coeff[i, 1])
}
"FINAL FOR THE THREE STATISTICS:"
"1: Posterior inclusion probability"
"2: Posterior mean conditional on inclusion"
"3: Posterior standard deviation conditional on inclusion"
final = (postprob, coeff, sd)
varnames
```

```
final
varnames

end
```

Appendix C

The following table contains the values reflected in Figure 23.

Table 22: Feasible long-range improvement trend

Year	Score	Year	Score	Year	Score	Year	Score
0	260						
1	270	41	492	81	556	121	589
2	280	42	495	82	557	122	589
3	289	43	497	83	558	123	590
4	299	44	499	84	559	124	591
5	308	45	502	85	560	125	591
6	317	46	504	86	561	126	592
7	325	47	506	87	562	127	593
8	334	48	508	88	563	128	593
9	342	49	510	89	564	129	594
10	350	50	512	90	565	130	595
11	357	51	514	91	566	131	595
12	364	52	516	92	567	132	596
13	371	53	518	93	568	133	597
14	378	54	520	94	568	134	597
15	385	55	522	95	569	135	598
16	391	56	523	96	570	136	599
17	397	57	525	97	571	137	599
18	403	58	527	98	572	138	600
19	409	59	528	99	573		
20	414	60	530	100	573		
21	419	61	531	101	574		
22	424	62	533	102	575		
23	429	63	534	103	576		
24	434	64	536	104	577		
25	438	65	537	105	577		
26	443	66	539	106	578		
27	447	67	540	107	579		
28	451	68	541	108	580		
29	455	69	542	109	580		
30	458	70	544	110	581		
31	462	71	545	111	582		
32	465	72	546	112	583		
33	469	73	547	113	583		
34	472	74	549	114	584		
35	475	75	550	115	585		
36	478	76	551	116	585		
37	481	77	552	117	586		
38	484	78	553	118	587		
39	487	79	554	119	587		
40	490	80	555	120	588		

Education and country growth models

Part II

Modelling optimal prioritisation across levels of education in the country development process

---

Martin Anders Gustafsson

CONTENTS FOR PART II

Figures and tables

7 INTRODUCTION FOR PART II

Part II of the dissertation addresses the following question: *What new patterns can be found in the historical country-level enrolment and per student spending data to inform policymakers on how to prioritise the various levels of the education system?* It will be argued below that given the importance of this question to policymakers, and the data that are available, not enough work in relation to the question has occurred. Some related modelling using country-level data has occurred and this will be discussed. However, what has been largely absent is rigorous analysis of a rather well-known data source, the data of the UNESCO Institute for Statistics (UIS), gathered over several decades from ministries of education around the world. The analysis presented below focuses above all on the enrolment and per student spending statistics, by education level, from this data source, going back to 1970. It moreover explores correlations between the education values and economic growth values.

Why is the question an important one? The circumstances of each country are different and there can clearly be no one-size-fits-all prescription relating to, for instance, what average public spending per student should be at the primary, secondary and tertiary levels for countries at different levels of development. Yet cross-country patterns can be useful insofar as they highlight when country-specific patterns are anomalous and might therefore be problematic and in need of an adjustment. The patterns can thus be suggestive. Currently the data analysis available to inform across-level prioritisation in education is sparse. The same can be said about the available theory. Policymakers have considerable decision-making power with respect to how public spending is spread across the levels of the education system. It is in this area that policymakers are likely to be particularly interested in analysis relating to the above question. Even with respect to the spread of enrolments across levels of the system, policymakers exert considerable influence, though arguably less so than in the case of public spending. Governments can increase or restrict the supply of student places at education institutions, though restrictions often result in serious public resistance and increasing supply may not result in actual enrolment increases if the demand for education is insufficient or there are quality bottlenecks, for instance insufficient graduates of the required standard exiting from a lower level.

Section 8 below proposes a set of factors which are likely to influence decisions relating to prioritisation across levels amongst policymakers. This examination includes some discussion of how policymaking processes work and how decisions are taken when there is little empirical evidence. It also includes discussion of how the microeconomic evidence on rates of return to education has to some extent been misinterpreted in the context of prioritisation across levels. Section 8 thus provides an institutionally-focussed examination of how decisions seem to be taken, which then forms a backdrop against which to explore, in the next section, more data-driven decision-making frameworks.

Section 9 looks at what the existing theory and empirical evidence in economics has to offer the policymaker faced with level prioritisation decisions. On the theory side, human capital theory, rates of return to education and models of country development receive attention. Modelling by Vandenbussche, Aghion and Meghir (2006) of OECD country data is examined, as is related work by Lutz, Cuaresma and Sanderson (2008) focussing on developing countries.

Section 10 begins by describing the UIS enrolment and expenditure data used for the new analysis, including the shortcomings of these data. Key shortcomings worth noting at this point already are the absence of reliable and comprehensive values on participation in or expenditure on pre-school and vocational training. The analysis thus focuses largely on relationships and trade-offs between the three large categories: the primary, secondary and tertiary levels. In fact, much of the emphasis in the analysis falls on the secondary and tertiary levels, as this is where policymakers have most discretion. Universal participation in primary schooling is generally considered an *a priori* policy priority, though expenditure per primary student is a more debatable matter. A discussion of the basic line of inquiry and the kinds of data analysis techniques employed follows. The analysis itself includes fairly basic descriptive analysis as well as some multivariate modelling. As will be seen, the data do not lend themselves to the identification of cause and effect. However, they do allow for the identification of associations which, it will be argued, are suggestive of actual historical dynamics and thus important for policymakers. It should be remembered that despite the fact that the UIS data are well-known amongst researchers and policymakers, they appear not to have been subjected to the kind of pattern-seeking analysis presented in section 10. Key findings include the fact that countries with faster growing economies have pursued different enrolment and education spending trajectories and, somewhat unexpectedly, a high positive correlation, even when controlling for a large range of non-education variables, between higher per student secondary school spending in the past and faster growth. In keeping the other two parts of the dissertation, the analysis focuses strongly on concerns that policymakers in developing countries would have, including policymakers in South Africa. Section 11 concludes Part II of the dissertation.

8 HOW EDUCATION POLICYMAKERS DEAL WITH PRIORITISATION ACROSS EDUCATION LEVELS

Education policymakers, and government policymakers in general, do to some extent actively influence how the various levels of the education system are prioritised, through funding decisions taken and the design of policies relating to matters such as entrance requirements, compulsory school attendance, and exit qualifications. At the same time, prioritisation across levels occurs partly through dynamic social and economic processes over which the policymaker has little direct influence. Below, six factors which are likely to influence how prioritisation occurs are discussed: available public funding; public demand for education; demands of employers; teacher union demands; international pressure; welfarist considerations. The first five factors are powerful external factors over which the policymaker is likely to have limited control, though how exactly these factors shape policies is still largely under the control of policymakers. The sixth factor is essentially economically-informed rational decision-making and is thus considered an optimal factor. The greater the degree to which prioritisation decisions are based on theories encompassing rational notions of efficiency, equity and country development, and the greater degree to which these decisions are based on empirical evidence, the better. The discussion of the six factors will point to certain complexities, including dynamics between the six factors and the fact that what may look like a rational pursuit of welfarist principles, may instead be something else.

(a) Available public funding. Public funding constraints limit the possible size of the education system as a whole. However, these constraints tend to be felt most acutely at the tertiary and to some extent secondary levels, because per student costs tend to be higher the higher the level (as seen in section 10). Moreover, any discussion of limiting in the size of the system due to funding constraints is likely to focus on, firstly, the tertiary level and, secondly, the secondary level, as limiting expansion at a higher level has only a limited effect on lower levels, whilst the reverse does not apply. In developed countries, it is the relatively high availability of public funds that has been a key factor in the achievement of extremely high enrolment ratios at all levels, making universal tertiary education appear as a possibility (Schofer and Meyer, 2005: 898).

(b) Public demand for education. Policymakers, in particular within democratic societies, are likely to be under pressure to expand access at whichever level universal coverage has not been reached to satisfy demands from citizens wishing to maximise their future incomes. At the primary level, but even at the secondary level, this pressure may take the form a constitutional provision making education a human right. However, not all popular pressure is oriented towards social equity. In less equal societies, an elite may create pressure for public funds in education to be largely directed towards this elite. Verspoor and Bregman (2007: 6), in a report produced for the World Bank, argue that this has been the case in Sub-Saharan Africa, insofar as elites have lobbied for the maintenance of elitist secondary school systems with high per student spending and relatively low coverage. One argument elites may use against a rapid expansion of the education system is that this reduces not just the average quality of graduates, but even the overall number of graduates achieving at a high level, through a dilution effect. The empirical evidence for such an effect is not strong, however, as argued in for instance Taylor and Spaull (2013).

(c) Demands of employers. Policymakers are influenced by the needs of employers, including the government as an employer, in a number of ways when plans are made to prioritise certain specialisations or expand particular levels of the education system. However, as Bertrand (2004: 71) argues, in a planning manual of the International Institute for Educational Planning (IIEP) directed at policymakers, labour market demands are too frequently ignored, or misinterpreted, in the education planning process. This is partly because, following the discrediting of highly detailed 'manpower planning' methods due to problems around the workability of these methods, the development of alternative techniques to promote labour market responsiveness amongst education planners has been weak. There has been, to some extent, a methodological vacuum. Yet there are large differences between the approaches of different governments, with some paying very little attention to labour market demands and others setting up sophisticated information and knowledge systems that at least improve the probability that education planners will understand what the labour market demands are. An example of the latter would be the Occupational Outlook Handbook (OOH) produced periodically by the Bureau of Labor and Statistics in the United States. The OOH includes a database, which can be queried online, covering over 1,000 occupations and providing statistics on projected growth in each occupation. For instance, 'high school teachers' require a Bachelors qualification and their demand is expected to increase by 7% between 2010 and 2020, the time horizon of the OOH forecasts (http://www.bls.gov/ooh/, accessed July 2013). The occupation 'construction labourers and helpers' requires no formal qualifications, though certain basic skills are needed, and here demand to 2020 is expected to increase by 23%. Clearly this information can guide education planners directly, though it is also intended to steer public demand for education. The United Kingdom's equivalent of the OOH, the *Working futures* report, also uses a ten year horizon but is more analytical, providing for instance forecasted trends with respect to the demand for graduates at different levels of the education system (United Kingdom: UK Commission for Employment and Skills, 2011: 120). However, Bertrand (2004: 44) suggests that even in the UK, the utilisation of this kind of information is limited. In South Africa, the Human Sciences Research Council (HSRC) has produced two issues of the *Human resources development review*, in 2003 and 2008, partly to provide research to education planners on labour market demands (Kraak and Press (eds.), 2008). There is little evidence that the HSRC's reports are used as such reports should be in education policymaking in South Africa. The reasons for this could of course be both a lack of responsiveness in the education sector and limitations of the reports themselves.

(d) Teacher union demands. Teacher unions demands for higher salaries at the primary and secondary school levels, where unionisation is typically strong and market forces weak, are likely to crowd out funding for post-school education, where unionisation tends to be relatively weak. In fact, evidence from several developing countries presented by Pritchet and Filmer (1997), and discussed in Part I of the dissertation, suggests that this kind of crowding out could be the norm as teachers mostly succeed in inflating their salaries to a point well above a more market-related level.

(e) International pressure. Part III of the dissertation discusses briefly the way foreign and international donors play a large role in the education decision-making of a specific set of developing countries. Researchers such as Samoff (1999) have been highly critical of this kind of pressure, including the way in which it has apparently led to an over-prioritisation of primary schooling at the expense of the higher levels of the system. The basic argument is that donors tend to over-simplify and misunderstand the dynamics of the recipient country and hence advocate and generally succeed in imposing inappropriate policies. How rates of return analyses have been misused, in particular by World Bank analysts advising Sub-Saharan African governments on how to prioritise education levels, is a matter that has received considerable attention in the literature. This is discussed towards the end of this section. As is shown in the data analysis of section 10, however, the movements in the key enrolment indicator values since 1970 do not clearly support the hypothesis of a perverse over-prioritisation of the primary level in the case of Sub-Saharan Africa.

(f) Welfarist considerations. Finally, more skilled policymakers take the five factors mentioned above and incorporate them into what can be called a welfare economics framework, even if not all policymakers would be happy to use this term. Essentially, this involves looking at what kind of prioritisation across the education levels would maximise future welfare, whilst recognising that optimal future welfare is to some degree subject to differing interpretations of social equity. Educational equity is likely to promote income and socio-economic equity, but at the same time using education to create a critical mass of talented experts, which inevitably involves allowing some inequity, can promote innovation and thus economic growth. Of course the analysis can occur in a more or less formal manner. Moreover, the use of empirical evidence is in itself not proof that good evidence-driven policymaking is occurring. The evidence should be correctly interpreted. It is not impossible that optimum decisions will be taken even in the absence of a rational planning approach characterised by sound interpretation of empirical evidence. Responding very directly to, for instance, the demands of the public may result in an optimal distribution of education funding across levels. However, it can be considered preferable to pursue optimum policies on the basis of a approach model, as opposed to hoping that optimum policies will be achieved by accident.

In any policymaking process, it is particularly important to avoid what is referred to as ‘piecemeal welfare economics’ by Lipsey and Lancaster (1956: 11), two economists who came up with the first formal definition of ‘second best’ policymaking. In this type of policymaking, one explicitly acknowledges that a Pareto-optimal economy may be rendered impossible by a constraint, generally a market imperfection, which is not removable. For instance, an optimal set of enrolment ratios and per student spending values may involve a tertiary enrolment ratio of x and tertiary per student funding of y. However, teacher union pressure may result in a crowding out of the tertiary level in the budgeting process and policymakers may respond by suppressing per student funding at the tertiary level to a point below y, whilst maintaining tertiary enrolments at x. Because of the constraint imposed by teacher unions at the primary and secondary levels, the optimal policy positions at the tertiary level no longer hold true. Either x or y, or both, stipulated in the ‘first best’ solution must be abandoned. Haphazardly reducing per student spending could represent ‘piecemeal welfare economics’. However, rationally determining how to achieve a sub-optimal deviation from both x and y would represent second best policymaking.

Turning to the problem of rates of return to education and their implications for prioritisation across education levels, in Part I arguments put forward by Glewwe (1996) were discussed briefly. His two key complaints around the way rates of return have been used are, firstly, that the exclusion of positive externalities within these rates results in an under-valuation of tertiary education and, secondly, that the years of schooling variable used in the calculation is a poor proxy for what people actually learn, meaning there is a serious and confounding measurement error in the calculation. Bennell (1996) offers further reasons for rejecting the way rates of return have been used to inform policy, in an article titled 'Using and abusing rates of return'. He argues that apart from possible problems with the source data, very basic anomalies in the rates of return calculations done by Psacharopoulos (2004) have been overlooked. He also argues that adjustments that ought to occur to typical rates of return statistics to take into account issues such as innate ability, socio-economic status and opportunity costs associated with informal sector work, would tend to reduce the primary level rates of return statistics to a larger degree than the tertiary level statistics. Gustafsson and Mabogoane (2012: 353) emphasise that rates of return, even in the absence of measurement errors, would tend to reflect the dynamics of a sub-optimal education system, and implicitly reject an approach where planners do not consider what the rates of return might have been in the presence of more efficient policies. Specifically, they link the high rates of return associated with attaining grade 12 in South Africa to the fact that grade 12 is where the schooling system's only national examination exists. The policy priority should perhaps rather be to introduce additional examinations lower down in the schooling system, as opposed to simply prioritising grade 12 attainment on the basis of existing rates of return.

There appears to be no published attempt to come up with a theory on how rates of return, with all their shortcomings, could be used to inform education policymaking, in particular prioritisation across different types and levels of education. Whether such a theory is possible and could assist policymakers is not clear. Any theory would need to take into consideration the key dynamics discussed above, such as positive externalities associated with higher levels of education. It would also need to move beyond simply asking which level of education to prioritise. It seems preferable to see the problem as one of an equilibrium. Enrolment growth and spending increases must probably occur across several levels of the system at once, so questions should focus on how to pursue an optimum spread of budgets and efforts across several competing priorities. The analysis provided in section 10 can be considered a contribution towards a more evidence-based and rational approach to prioritisation across education levels, even if it falls several steps short of comprising a neat and comprehensive theory.

9 THE ECONOMICS OF PRIORITISING ACROSS EDUCATION LEVELS

9.1 The basic economic theory

Figure 36 below presents a diagram depicting the theoretical framework for investment in education by households and the government. The framework represents a fairly standard picture of the human capital model, with some differentiation by education level added (Schultz, 1961; Boissiere, 2004). A diagram such as this could be of use for education policymakers wanting to organise the discussion around the prioritisation of the levels of the education system, where the ultimate aim is to advance economic growth, or national development. The diagram can also be considered a summary of various strands of theory that are discussed across the three parts of the dissertation.

A useful point of departure is the decisions taken by the household to invest in the education of its younger members. Investment decisions here are influenced by a number of variables. The household perceives, perhaps inaccurately, what incomes are associated with specific years of schooling and qualifications. These perceptions can encourage investment in more education, depending on the income gained from the additional education relative to the income forfeited from not entering the labour market immediately. The latter is often referred to as the indirect or opportunity cost of education. High direct costs, in the form of for instance tuition fees, could dissuade the household from investing further in education. The decisions of the household carry considerable risk, for two key reasons: it is risky to use present information as a proxy for what future incomes will be and the household cannot know if its members will succeed in reaching achievement levels required for specific qualifications. Investment in primary and secondary schooling often occurs due to the option value of this level of schooling, or the fact that it makes higher levels of education an option at a later point in time[1].

The education system gives household members two key things: learning outcomes, in the form of skills and knowledge, and qualifications. Importantly, learning outcomes are brought about not just by what happens in the education system, but also by what happens in the household. The greater the schooling, or human capital, of the previous generation of scholars in the household, or the parents, the greater the contribution of the household towards the learning of the next generation. Education makes people more productive, which contributes to higher incomes. However, success in the labour market also depends to some degree on qualifications, which provide signals to employers, and allow employers to screen job seekers. In theory, qualifications would reflect the skills and level of productivity of job seekers, though in practice the signalling devices would be imperfect. For instance, people with the same qualification would possess different innate abilities. It is possible for less qualified people to be more productive than more qualified people if, for instance, the qualifications system itself is inconsistent or unqualified people have acquired skills outside the formal education system.

[1] The term ‘option value’ is typically used to explain why governments invest in or maintain environmental or public assets which are not needed currently, but which are likely to be needed in future. McMahon (1982) and Burger and Teal (2013) are examples of economists who have used the term in the context of education investment choices.

Economic growth and social development occur as aggregate income rises, but also as positive externalities arising out of a more educated society, for instance greater trust and social cohesion, are realised. There is a cycle of improvement as each generation uses some of its income and capabilities to ensure that the next generation attains an even better level of education. A key question is how governments and policy can accelerate this cycle of development, or perhaps undermine it. This is dealt with in the grey boxes of the diagram.

Government can accelerate human capital accumulation by ensuring that the redistribution of income, through the tax system, manifests itself in subsidised education for the poor. This redistribution plays a particularly large role in promoting growth in the lower levels of the education system, specifically primary schooling. Virtually everyone, regardless of socio-economic status, is able to participate in primary schooling. In contrast, home background disadvantages are a key reason why many are not able to fulfil entry requirements for tertiary education institutions. Moreover, due to high private returns to tertiary education, private loans for this level of education are, at least in theory, relatively easy to obtain, even for the poor who have performed well at the secondary level. In reality, however, poorly functioning banking sectors in developing countries may make it difficult for these loans to be realised. The fact that the cost of educating each primary school child is relatively low, is a further motive for a government to subsidise, often fully, primary school education for all. In fact, governments will often subsidise even the primary schooling of the rich, who could afford to pay for this schooling through private means. There is no strong economic rationale for any government to do this, but an ideological concern to keep all of society within one public primary system is often the determining factor here. Governments will often enforce investment in primary schooling through compulsory education laws. The enforcement applies particularly in the case of the very poorest households, which might otherwise have benefitted economically from having their young children work for an income, or in a family business or farm. Again, the overriding influence here may be ideological, rather than purely economic. Primary schooling has come to be seen as a non-negotiable basic human right or, to use the more economic terminology, a merit good.

Due to the very public nature of the education system, prices are often not determined through market forces, but administratively, which increases the risk of rent-seeking (capture of income beyond what can be justified by one's productivity) and inefficiencies. A large policy challenge revolves around the determination of teacher salaries in the pre-tertiary schooling sector, whose structure lends itself to high levels of unionisation and thus a combination of monopoly (a union will represent all teachers) and monopsony (the state is the only buyer of services). The government needs to ensure that the monopolistic result of inflated prices (or salaries) or the monopsonistic result of deflated prices are avoided. This needs to be achieved to a large extent through an analytical process of comparing teacher productivity and the wages of teachers relative to those of other similar professionals. These kinds of problems are relatively absent at the post-school level, largely because university teachers tend to have skills which are in demand in various sectors of the economy, thus allowing for roughly a market-like determination of wages.

Governments typically need to find innovative ways of enhancing productivity in the education sector, given the serious information asymmetry problems that are inherent to the education sector. The clients, households, are often unable to understand the quality of the specific education they are investing in until it is too late, in other words when the scholar enters the labour market. Even the government experiences great difficulties in assessing whether learning outcomes could be increased with the given budgets and technologies. Accountability systems are needed, but these should be designed in such a way that they minimise the possibility of gaming, in particular the possibility that teachers will misinform the administration about performance in their institutions. Incentives and disincentives can be attached to accountability systems, but unless this is done carefully, it can encourage even more misinformation. Moreover, poorly designed incentives can create unintended disincentives if, for instance, they result in a sense that the administration treats teachers unfairly.

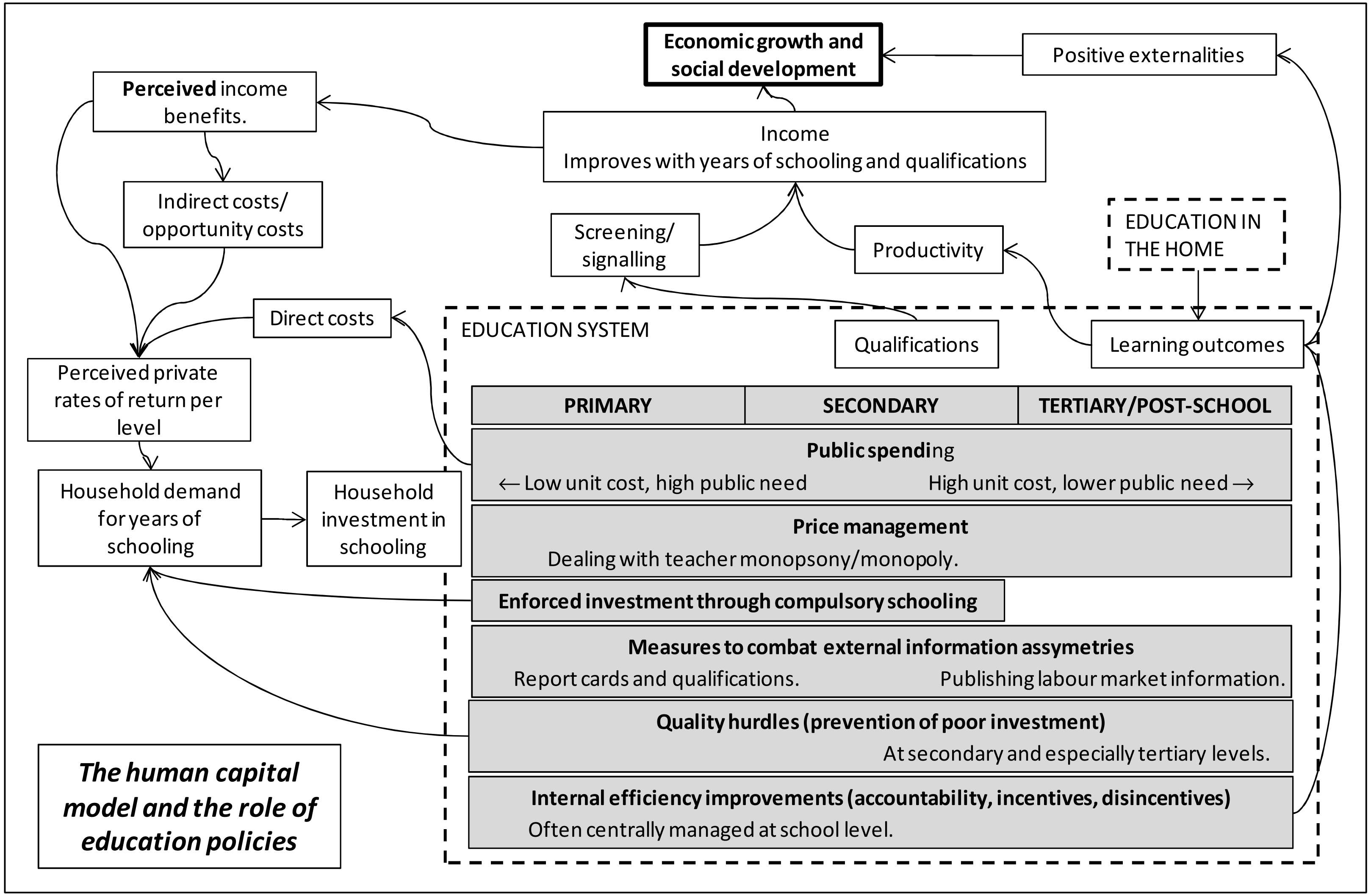
The government can reduce the information asymmetries experienced by the household in two important ways. Firstly, it can ensure that the household receives good information about the educational performance of its scholars, through standardised assessment systems. Not only does this assist the household in determining whether it is worth educating the scholar further, this information also allows the household to become an agent in the education process. An informed household is able to complain about the quality of the education service, may be in a position to consider alternative schools and might, together with other households, exert political pressure on the school to perform better.

Secondly, the government can make it clearer to the household what level of income is associated with different types of education, especially at the post-school level, where the heterogeneity of educational offerings can be confusing. It can do this by collecting data and publishing the average earnings of graduates from different fields of study and institutions.

Just as the government can enforce investment in primary schooling in the interests of combating poverty and advancing economic growth, it can also prohibit the household from investing in education where this seems like a poor investment. It does this through a system of quality hurdles, often in the form of examinations, which prevent scholars from advancing further in the education system if they are considered unable to succeed at the next level, or within a specific field of study.

What has been provided here is a high-level summary of key dynamics that should influence the education planner's thinking around the prioritisation of the various levels of the education system. Clearly, the amount of education that happens at each level, in terms of enrolments and learning outcomes, is not a simple matter of spreading budgets across the levels in a manner that is appropriate for the country's level of economic development. Various interventions aimed at improving the education system's internal efficiency, setting standards and making information available interact with each other and determine the effectiveness of each level, and of the of education system as a whole.

Figure 36: The human capital model and the role of education policies



9.2 Empirical findings from cross-country models

As discussed in section 8, the use of the microeconomic rates of return to education evidence to inform prioritisation across levels of the education system has been flawed. It is not clear whether the existing flaws in this channel between evidence and policymaking can be corrected, but this is not the focus of this dissertation. Instead, the focus in the dissertation is mainly on the use of cross-country data in answering policy questions. Part II pays special attention to how these data can be used to solve policy questions relating to the prioritisation of education levels. Here there seems not to be much empirical work done beyond the few texts discussed in this section. Two of these texts are discussed in some detail as their coverage of the policy questions is rather comprehensive: Vandenbussche, Aghion and Meghir (2006), who focus just on high income countries, and Lutz, Cuaresma and Sanderson (2008), whose focus is developing countries. At the end of the section, a few further texts are briefly discussed which deal less comprehensively with the question posed at the start of section 7, or limit their attention to specific geographical regions.

Vandenbussche, Aghion and Meghir (2006), henceforth referred to as VAM, present a series of growth regressions which are unusual insofar as they focus on optimal mixes of levels of education in the labour force. Their theoretical framework was discussed in Part I of the dissertation. Here their empirical work is discussed. VAM are largely motivated by a need to respond to Krueger and Lindahl (2001: 1119, 1130), who found that average years of schooling of adults was not correlated with economic growth in models covering just high income countries, and hence concluded that education was a less important driver of growth in these countries than was commonly thought. This explains partly why VAM use only data from high income countries, specifically 19 high income OECD countries. Further reasons for this are the limited nature of certain data from developing countries, in particular historical public expenditure data and average years of schooling attained data, and the fact that VAM were interested in examining trade-offs between imitation and innovation, the second of these phenomena being a feature largely of developed countries. Despite the exclusion of developing countries, the analysis presented by VAM would probably be of interest to developing country policymakers, who are often under pressure to pursue innovation, often on the basis of ideological considerations.

A key concept in VAM is distance from or proximity to the global technological frontier. If their model is important for education planning, and they suggest it is, then so is understanding a country's proximity to the technological frontier. They hypothesise that how close a country is to having internalised all the best available technology at any point in time determines what the optimal split in the labour force between skilled and unskilled workers is. In their empirical analysis, VAM consider workers with post-school education as skilled and those with just school education as unskilled. They furthermore assume that the technology of the United States represents the technological frontier. Distance from this frontier is measured in terms of total factor productivity (TFP), a measure worth discussing in some detail. TFP is essentially a measure of how productive workers and the technologies they use are. In other words, it can be seen as the *A* of the Cobb-Douglas model discussed in Part I of the dissertation. VAM (2006: 110) define the term as follows: 'output per adult minus capital per adult times the capital share'. This means that TFP is higher when (1) more adults are workers, (2) workers are more productive given a particular level of technology, (3) the technology used is more advanced and (4) the capital share is higher. The capital share and labour share of an economy together equal one. The capital share is generally calculated as the part of national income that is not paid in the form of wages but instead takes the form of interest income and economic profit (Schneider, 2011: 1). TFP improves when the capital share is higher, if output remains the constant, as this would mean that the ratio of outputs to wages increases. VAM themselves do not elaborate on issues relating to the measurement of TFP, though a discussion in Manuelli and Seshadri (2010) suggests this is not a straightforward matter. Saliola and Seker (2011) use firm data to calculate TFP and find within a group of 27 developing countries Zambia, South Africa and Colombia to be amongst the bottom five countries and Ethiopia, Mexico and Peru to be the top three. Intuitively these rankings are difficult to grasp, suggesting that TFP may not have become a sufficiently robust measure to inform policymakers.

Notwithstanding some doubts around measuring proximity to the technological frontier, as a whole the VAM model does appear to make intuitive sense. In one key growth regression (VAM, 2006: 113, model 5), data from five-year intervals for the period 1960 to 2000, covering the 19 OECD countries, are used in a regression described by the following equation:

$$\begin{aligned}(\ln Y_{it} - \ln Y_{i,t-1}) = B_{0,j} + B_{1,t} + B_2(\ln Y_{i,t-1} - \ln Y_{*,t-1}) + B_3 h_{i,t-1} \\ + B_4((\ln Y_{i,t-1} - \ln Y_{*,t-1})h_{i,t-1}) + u_{it}\end{aligned} \tag{86}$$

On the left-hand side is growth over a five-year period with respect to *Y,* which in this equation is taken to be total factor productivity. On the right-hand side, fixed effects are introduced for each of seven groups of countries, determined according to institutional and geographical criteria, and indicated by the subscript *j*, as well as for each time period *t*. The first slope coefficient describes the association for distance to the technological frontier, lagged by one period, where * indicates that the TFP of the country with the maximum value in period *t* is used. The variable *h* is the percentage of adults who have attained a tertiary education. This is lagged by one period. Finally, the last explanatory variable is the interaction of the previous two. All three coefficients B_2, B_3 and B_4 are statistically significant at least at the 5% level. What is not indicated above are three instrumental variables: distance to the frontier lagged by ten years, an indicator of expenditure on tertiary education lagged by five years, and the interaction of these two. The last two of these variables emerge as statistically significant in the first stage regression. VAM conclude, on the basis of a rank test, that endogeneity, specifically the possibility that improvements in *Y* cause improvements in tertiary attainments, is not a factor that needs to influence the interpretation of the results.

A simulation using the results of the regression indicate that if a country that has incorporated 80% of the technology of the frontier (amongst the 19 high income countries the mean is 74% and minimum 42%) and has achieved 13% attainment of tertiary education amongst adults (mean of 13%, minimum of 1.5%) increases the attainment percentage from 13% to 15%, annual growth improves by 0.1 percentage points. However, had the country incorporated 85% of the technology of the frontier, the same improvement to the attainment percentage would result in an improvement to annual growth of a much larger 0.25 percentage points. VAM (2006: 122) believe the results have policy relevance: 'Our empirical estimates in this paper can be used to derive policy prescriptions in developed countries'. How exactly this might occur is debatable, but what is clear is that planners stand to gain from improvements to the measurement of the distance to the global technology frontier, as this statistic influences assessments such as whether too few graduates are being produced by the tertiary education sector. The analysis by VAM provides empirical backing for the intuition of education planners, specifically the notion that expanding the tertiary education sector is likely to impact positively on growth, though, as shown by VAM, there is an important conditionality attached to the expected magnitude of the growth improvement. The weakness with the Krueger and Lindahl (2001) argument is thus that their measure of educational attainment was just one average value encompassing all levels of education, and not level-specific values. Disaggregation by level is necessary if the relationship between education and economic growth is to be properly understood. Of course further analysis along the lines of VAM to explore the dynamics in developing countries would be valuable.

In Part I of the dissertation, the IIASA-VID dataset of internationally comparable educational attainment statistics by age, developed jointly by the International Institute for Applied Systems Analysis (IIASA) and the Vienna Institute of Demography (VID) was described. Lutz, Cuaresma and Sanderson (2008) describe a growth regression that uses these data and that probably comes closer than any other available analysis to answering empirically the question of how growth at the various levels of the education system should be prioritised in the interests of country development and economic growth[2]. The analysis has certain drawbacks, which are discussed below, yet it stands out as filling a gap in the body of research required by education planners. Perhaps because the research is still relatively new, but perhaps also due to the unresponsiveness of policymakers, there seems to be little evidence that this research by Lutz *et al* (2008) has been used or critiqued by those engaged in policymaking.

The growth regression by Lutz *et al* (2008) uses data from 101 countries, 84 of which are non-OECD countries, for five-year intervals in the period 1970 to 2000. The empirical form of the regression, seen below, is in several respects similar to that of VAM.

$$\begin{aligned}(\ln Y_{it} - \ln Y_{i,t-1}) = B_0 + B_{1,jk} \sum_j \sum_k (\ln L_{ijkt} - \ln L_{ijk,t-1}) + B_2 (\ln W_* - \ln W_{it}) \\ + B_3 (\ln K_{it} - \ln K_{i,t-1}) + B_4 \sum_j \sum_k \left(\frac{L_{ijkt}}{L_{it}} \right) (\ln W_* - \ln W_{it}) + u_{it}\end{aligned} \qquad (87)$$

Here Y is GDP in real terms, so the left-hand side of the equation refers to GDP growth for country i over a five-year period. On the right-hand side, there is one undifferentiated intercept, so no fixed effects are employed. The first log difference on the right-hand side is repeated eight times, for two broad age categories j, being 15 to 39 and 40 to 65, and four educational attainment categories k, being: (1) no formal education, (2) from incomplete primary schooling to incomplete lower secondary schooling (called 'Primary' in Table 23 below and in the original research article), (3) from complete lower secondary schooling to incomplete first level tertiary education (called 'Secondary' in the table), and (4) first level of tertiary complete and above

[2] The details of the growth regression appear in a separate technical appendix.

(called 'Tertiary' in the table). *L* is the number of adults in each of the eight categories, so the explanatory variables deal with growth in a country's human capital with respect to the two age categories and the four education attainment levels. The equation includes a distance to global technological frontier variable, as in equation (86), except here the measure is simply the gap between the per capita income, *W*, of country *i* in period *t* and the highest per capita income for any country or period found in the dataset. Growth in the capital stock *K* is the next explanatory variable. Finally, a series of variables dealing with the interaction between level of *L* and distance to the global frontier is included.

Of special interest to policymakers would be the simulations, presented by Lutz *et al* (2008: 1048), for four hypothetical countries, the results of which are reproduced in Table 23. The expected annual growth associated with four different combinations of human capital amongst adults is described.

Table 23: Simulations from regression using IIASA-VID data

Country	No schooling	Highest level attained			Total	Annual growth
		Primary	Secondary	Tertiary		
A	50	40	10	0	100	2.0
B	0	90	10	0	100	6.5
C	0	50	50	0	100	13.0
D	50	30	15	5	100	7.0

Source: Values read off graph in Lutz, Cuaresma and Sanderson (2008: 1048).
Note: See preceding discussion for definitions of the education levels.

As argued by Lutz *et al* (2008), what seems particularly relevant for policymaking is country C, which illustrates the especially large returns associated with investments in secondary schooling, remembering that this has been defined as ranging from attainment of complete lower secondary schooling to incomplete first level tertiary education. It is implied that the United Nations, with its strong prioritisation of universal primary education (UPE), has probably under-estimated the importance of secondary schooling. Specifically, if moving from the country B scenario of universal primary education with very little secondary education to the country C scenario of 50% attainment of secondary schooling should bring with it additional growth returns, of 6.5 percentage points, which are considerably larger than the returns associated with a movement from the country A scenario, of very low education attainment generally, to the UPE of country B, then the general perception amongst many education planners and donor agencies has probably been somewhat misguided. Country D challenges the common opinion to an even greater degree. It implies that not insisting on UPE and maintaining a relatively elitist education system results in a level of growth which is at least as good as the UPE with a bit of secondary schooling scenario of country B. This could be of interest to planners in countries that by 2010 were still far from achieving UPE. The IIASA-VID data point to the existence of six such countries in 2010, all them in Sub-Saharan Africa (http://webarchive.iiasa.ac.at/Research/POP/edu07/, accessed November 2012). Of course the notion that basic education is important as a human right, and not just as human capital investment, would disqualify the country D option. What seems of greatest relevance for policymakers in Lutz *et al* (2008) seems to be the evidence pointing to the need for a strong focus on secondary schooling in developing countries. Above all, in the absence of strong policy signals around secondary schooling, there is a risk that once developing countries attain universal primary schooling, there will be a sense that 'the job is done' and public resources will not be directed towards secondary schooling to the degree they should be.

One drawback with Lutz *et al* (2008) is that they do not consider the possibility of endogeneity in their growth regression, as VAM do. In particular, in drawing conclusions around the impact of investment in the various levels of the education system, they do not consider the possibility that in the past economic growth has driven investments in education, as opposed to the reverse being the case. However, the direction of causation in this relationship has been studied extensively by other analysts, including VAM (2008) and researchers such as Hanushek and Woessman (2009) discussed in Part I of the dissertation. The evidence suggests strongly that the direction of causation is largely from education to growth.

Turning to a few other attempts to answer the levels prioritisation question using country-level data, Mingat and Tan (1998) provide a series of essentially bivariate and cross-sectional analyses of a number of development and education indicators in an attempt to uncover the 'mechanics of progress in education'. Indicators cover, in particular, enrolment, grade repetition, school life expectancy, spending per student, spending on non-personnel, teacher pay, test scores, and demographic structure. A key concern in Mingat and Tan (1998) is how factors such as a decrease in teacher pay relative to GDP per capita allow richer countries to devote more resources to each pupil. The prioritisation of levels is a minor concern, though with progress is said to come a reduction in inequality between the per student spending levels of the secondary and tertiary levels. Yet the authors find considerable randomness in the data, which precludes the identification of clear development paths. Essentially, 'countries have substantial leeway to set priorities' (Mingat and Tan, 1998: 7).

Different ways of prioritising education levels, and types, are said to lie behind the slower growth of European countries relative to the United States in recent decades. Aghion and Howitt (2009: 287, 312), in reviewing the available literature, argue that lower levels of attainment of tertiary education in Europe relative to the US explain the higher growth of the latter. Krueger and Kumar (2004), relying almost exclusively on a theoretical and mathematical macro model, conclude that Europe has forfeited growth in the period since around 1980, relative to the US, because Europe did not respond to the acceleration of technological change by shifting priorities away from vocational education, which was an appropriate priority up to the 1970s when technological change was slower, and towards general education. In a context of rapid technological change, it becomes less feasible to train people in the use of specific technologies because that training soon becomes redundant. Instead, one needs to focus on giving people the fundamental cognitive skills they need to adapt to changing circumstances. Such skills are best developed through a more general education.

Teal (2011) uses evidence from a country-level panel dataset to argue that Sub-Saharan African countries should perhaps place an even greater relative prioritisation on tertiary education than other countries. In a regression analysis predicting income per capita for 32 African countries for the 1960 to 2004 period, Teal (2011: 58) finds that the average years of tertiary education, and to some extent secondary education, of adults are a far better predictor of growth than years of primary education. In some respects, Teal's (2011) policy recommendations are diametrically opposed to the notion, already discredited by others as seen in section 8 above, that Africa should prioritise primary schooling. Caveats put forward by Teal (2011) include the conditions that growth in the number of tertiary education graduates should be coupled to a shift towards more private sector employment and a stronger emphasis on the export of goods and services produced by graduates.

10 NEW FINDINGS FROM THE COUNTRY-LEVEL DATA

10.1 The data

The key data used for the analysis that follows are, firstly, income per capita from the Penn World Table version 7.0 (Heston, Summers and Aten, 2011) and, secondly, eleven indicators relating to enrolments and expenditure drawn from the UNESCO Institute for Statistics (UIS) online query facility (UNESCO: UIS). The eleven indicators are listed in Table 25 below.

Part I of the dissertation described a number of different UNESCO datasets, including the online UIS data on education, most of which are derived from questionnaires completed by ministries of education around the world. Part I also explained how the UIS data have been used, in combination with other data sources, to produce country-level datasets on educational attainment. Here issues relating to the eleven selected indicators are discussed.

The online UIS database is large and includes over 800 variables. The data used for the analysis were downloaded in September 2012. Only portions of the data can be downloaded at a time, so only data that appeared useful for the analysis were accessed. In fact, the labour-intensity of the process of downloading and then normalising the UIS data is likely to be one reason why these data have not often been systematically analysed. Another reason could be the limited technical documentation explaining the data. Obviously, the UIS database is expanded over time as new questionnaires covering new years are received, but retrospective changes to values seem to be rare, apart from changes relating to countries that have broken up, as discussed below. In the data used for the analysis, values were available for the 1970 to 2011 period, with at least some data being available every year. There were at least some values for 197 different countries. In selecting data, the completeness of the data across years and countries was one consideration, as well as relevance to the issue of level prioritisation. The UIS data follow the International Standard Classification of Education (ISCED), determined by UNESCO, in specifying levels of the education system. The following seven ISCED levels, from 0 to 6, exist (UNESCO, 1997).

Table 24: ISCED classifications

ISCED level	Description
0	Pre-primary education
1	Primary education or first stage of basic education
2	Lower secondary or second stage of basic education
3	(Upper) secondary education
4	Post-secondary non-tertiary education
5	First stage of tertiary education
6	Second stage of tertiary education

Levels 2 to 5 have A and B sub-categories, covering more general and more vocational education respectively.

Five gross enrolment ratios (GERs) are included amongst the eleven indicators, corresponding to the primary, lower secondary, upper secondary, secondary, and tertiary levels. These GER values are calculated by UIS. Lower and upper secondary GERs are disaggregated versions of the secondary level GER. The secondary level GER values cover all programmes, so vocational education at this level would be included. As seen in Table 25, lower and upper secondary GER values span a shorter period, from 1998 to 2011. No already calculated GER for ISCED level 4 is available, but a more or less equivalent indicator was derived, as explained below.

Table 25: Data availability

Variable	Values	Countries	Min. Year	Max. year
GER primary	6,089	193	1970	2011
GER lower secondary	1,979	190	1998	2011
GER upper secondary	1,885	189	1998	2011
GER secondary	5,303	193	1970	2011
GER tertiary	4,429	190	1970	2011
Vocational ISCED 2/3	4,605	189	1970	2011
Post-secondary non-tertiary	804	110	1998	2011
Tertiary engineering	824	132	1998	2011
Per student expenditure primary	2,187	165	1971	2011
Per student expenditure secondary	2,117	169	1971	2011
Per student expenditure tertiary	2,089	164	1971	2010
GDP per capita PPP in 2005 USD	6,891	189	1970	2009

The gross enrolment ratio is the enrolment of a particular level of the system, divided by the population of the age that would be enrolled at that level if there was no grade repetition and everyone in the population completed that level. Clearly, the GER can exceed 100 (the ratio is generally multiplied by 100). As already mentioned, the documentation accompanying the UIS data is thin, so a number of key criteria are not completely clear. For instance, there is some ambiguity around what is considered 'tertiary' in the UIS data. More critically, the age range in the population denominator used for the tertiary GER for each country is not made clear. The only specification that could be found was one on the World Bank site indicating that for all countries the five age cohorts starting from the age cohort immediately after the last secondary schooling age cohort, are used[3]. Given the close collaboration between UNESCO and the World Bank when it comes to education data, and the fact that the World Bank has no parallel data collection process for national education statistics but relies on UNESCO, the specification on the World Bank website is probably accurate.

Competing, in a sense, with the GER is the net enrolment ratio (NER) and, available since 2009, the adjusted NER. The NER is enrolment within an education level over population counting only those who are in the correct age range in both the numerator and the denominator. The NER should thus never exceed 100. The adjusted NER, which should also never exceed 100, is essentially an expanded age-specific enrolment ratio (ASER), for instance all 7 to 15 year olds enrolled at *any* level of the system, divided by all 7 to 15 year olds in the population. As pointed out by Lewin (2011: 7), the advantage of the NER over the GER is that it is much less likely to be distorted by inefficiencies in the form of grade repetition. One can illustrate the problem with an extreme example. If in one country only half of all children get to attend primary schooling, at the right ages, but all repeat each grade once, the GER

[3] http://data.worldbank.org/indicator/SE.TER.ENRR.

for the primary level would be 100, but the NER would be a more realistic 50. The disadvantage with the NER is that it is often distorted by poor data, given its reliance on what are often problematic age breakdowns in the enrolment data. Gustafsson (2012b) explains how even the GER is subject to a number of measurement problems relating to discrepancies between enrolment and population data. The NER would be subject to these same problems, but also further problems in the enrolment data. For many countries the NER is simply not calculated, whilst the GER is. In the 1970 to 2011 period, there are twice as many GER values as NER values in the UIS data. In the 2000 to 2011 period, 198 countries have GER values, but only 162 have NER values. Countries without NER values in this later period include India and Russia. Conceptually, the GER has some advantages. It does provide a sense of the capacity of the education system. Even if in our hypothetical country only half of all children receive primary schooling, if the GER is 100 then the country has enough capacity, in terms of physical space and teachers, to provide primary schooling for all, if only repeaters were replaced by non-repeaters.

Apart from the five GERs listed in Table 25, three new enrolment ratios were derived from the available UIS data to deal with important pockets of the education system. Firstly, technical and vocational education enrolments in ISCED 2 and 3 were divided by total ISCED 2 and 3 enrolments. Secondly, enrolments in something the UIS refers to as ‘post-secondary non-tertiary’ was expressed as a percentage of itself plus all tertiary enrolments. Thirdly, tertiary enrolments in ‘engineering, manufacturing and construction’ were divided by all tertiary enrolments, given the importance of these specialisations for economic growth and given that this new indicator displayed some interesting initial patterns. The Table 25 figures point to the fact that the second and third indicators described here could be calculated only for a limited number of years and countries.

Turning to expenditure by level, three indicators already calculated by the UIS were used. These were expenditure per student expressed as a ratio of GDP per capita, for the primary, secondary and tertiary levels. The available metadata make it clear that the numerator in this ratio is public expenditure, excluding private expenditure, and it is fairly clear that only current expenditure, and not capital expenditure, is considered (see for instance UNESCO, 2003). The questionnaires used by the UIS to collect data from ministries of education seem to suggest that respondents should fill in only spending directed at education institutions themselves, and not, for instance, the education administration outside these institutions. However, this is not altogether clear and in many education systems accounting systems are not able to distinguish clearly between, say, staff based permanently at schools and staff who move between schools providing administrative or educational support.

Only 197 countries were included in the analysis, and these were all countries listed in the *World Development Report* of 2012 (World Bank, 2011b). Very small countries not on the list were thus filtered out. With regard to country breakups and mergers, it should be noted that the UIS data include historical values from before the 1990s for many states created post-1990 in Eastern Europe and Central Asia. For instance, values for the Czech Republic and Slovakia going back as far as 1971 and 1987 respectively are available, though these countries only came into being in 1993 after the breakup of Czechoslovakia. What the UIS has not done is simply use the values submitted by Czechoslovakia for pre-1993 data on the two new countries. The pre-

1993 values are different for the Czech Republic and Slovakia. The methodology followed by the UIS is not clear. All 197 countries and details on their data availability with respect to the eleven education indicators appear in Appendix A of Part II of the dissertation.

10.2 Methodology

In proceeding with the data analysis that follows, it was assumed, correctly it seemed in the end, that a systematic exploration of patterns within the data summarised in Table 25 would produce findings that are relevant for policymakers, could inform future research and are interesting from a historical perspective. The analysis presented below employs an eclectic mix of techniques, including factor analysis, a semi-Bayesian growth regression model discussed in Part I of the dissertation and techniques commonly used in econometrics, in particular panel data fixed effects modelling. There is a strong emphasis on innovative ways of graphing the data and, implicitly, on constructing economical graphs that on the one hand convey as much relevant information as possible, whilst minimising the risk of being overly complex. The emphasis on graphs should be understood in the context of a key concern of the dissertation, which is to convey information to education planners as effectively as possible. There is some published literature supporting the efficacy of graphs for communicating quantitative findings, such as Kastellec and Leoni (2007). Probably the most widely known recent example of innovative graph use to influence a range of decision-makers is the work of Hans Rosling, often involving animated graphs. Whilst no animated graphs were developed in the analysis that follows, it is to some extent inspired by Rosling's work (Rosling, Rosling Rönnlund and Rosling, 2005).

The overriding research question for Part II of the dissertation, introduced in section 7 above, informed four data analysis questions, which are each dealt with in their own section below. Firstly, there was the question of whether a simple cross-sectional analysis, using just recent data, would reveal interesting patterns, along the lines of the work of Mingat and Tan (2003). Secondly, what patterns would the historical data display, and to what extent would these patterns support the more simplistic cross-sectional approach to understanding country development? Thirdly, are there correlations between economic growth and the various education indicators which reveal anything new, even if just suggestive, about the role of education in accelerating development? Fourthly, how could one practically use the patterns found in the data to inform policymaking in a specific country?

10.3 Cross-sectional analysis of recent values

Table 26 below offers a useful first view of the cross-sectional relationships between the eleven indicators listed in Table 25 above, on the one hand, and per capita income (in PPP terms) in 2009, on the other, on the basis of purely bivariate regressions. Income is expressed as log to the base 2, so coefficients point to the change in the indicator value associated with a doubling of income. To pick out a rather striking example, increasing the secondary level gross enrolment ratio (GER) by 14 points is associated with a doubling of income, and this education indicator is able to explain 61% of the variation in the income variable. Amongst the enrolment variables, the primary GER displays a poor correlation with income, largely because all countries have achieved relatively high levels of primary enrolment and poor countries tend to have more grade repeaters. A poor correlation is also observed for the three variables dealing with post-secondary non-tertiary, vocational and engineering enrolments. These three last variables will be discussed mainly within sections 10.4 and 10.5.

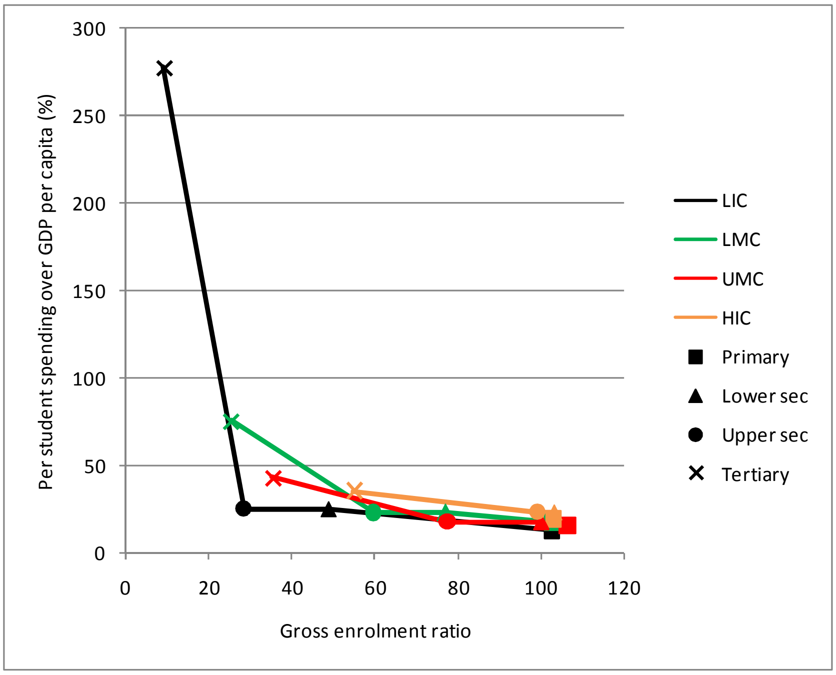
Table 26: Regression of indicator values on income

Variable	Countries	Coefficient on log of income to base 2	R^2
GER primary	183	1.2	0.023
GER lower secondary	182	12.0	0.608
GER upper secondary	181	12.9	0.663
GER secondary	181	14.4	0.614
GER tertiary	173	8.5	0.443
Vocational ISCED 2/3	161	2.3	0.138
Post-secondary non-tertiary	104	-1.9	0.039
Tertiary engineering	123	1.0	0.084
Per student expenditure primary*	144	0.8	0.042
Per student expenditure secondary*	141	-1.1	0.027
Per student expenditure tertiary*	132	-51.8	0.177

*Note: * all per student expenditure values expressed as a ratio of GDP per capita in 2009. Each row in the table refers to a separate two-variable regression.*

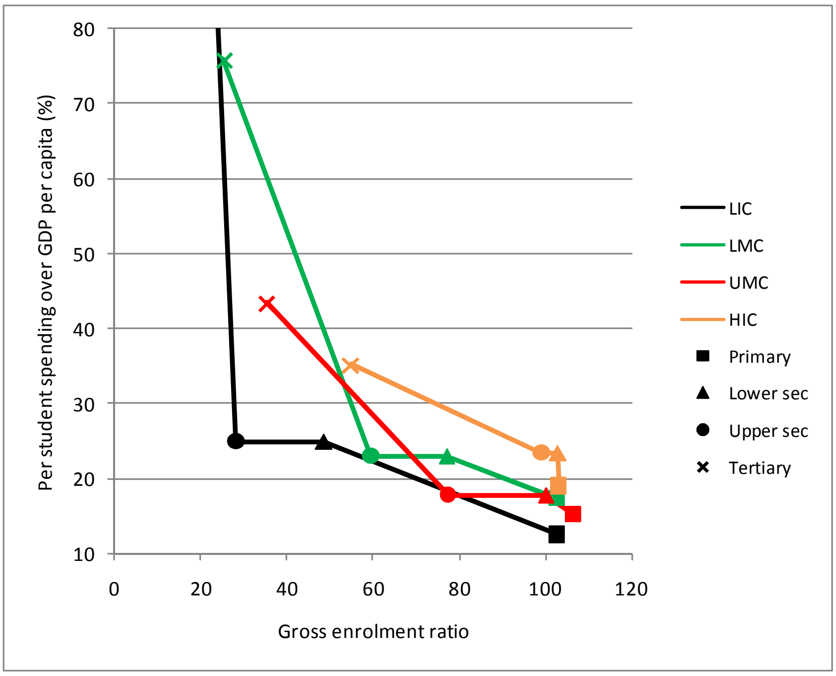
Given the kinds of concerns that education policymakers have, it seemed worthwhile to come up with a graph that could effectively display, for several countries or several groups of countries, enrolment and spending patterns across the three main levels of the education system. For this purpose, the format shown in Figure , with its magnification in Figure 38, was developed. Each curve displays where the means for the GER and expenditure indicators are situated, at the primary, secondary and tertiary levels, by four levels of country development: low, lower middle, upper middle and high income. For enrolments, secondary schooling was broken down into lower secondary and upper secondary as the data allowed for this, though not for a similar breakdown when it came to spending. Income classifications are those of the World Bank (2011b). To deal with the problem of missing values, in the case of each country data from the last year after 1999 with the lowest extent of missing values were used. Data from 194 countries are used for the graphs, the average number of non-missing values per country being 5.9, out of a possible maximum of 7 (3 enrolment values, 4 spending values). For 112 countries, all 7 values were present. The mean year of the country data used was 2007.0.

Figure 37: Educational level prioritisation ±2007



Source: Own calculations from dataset described in section 10.1. Values from 194 countries are used. Mean year across countries is 2007.0.

Figure 38: Magnification of Figure



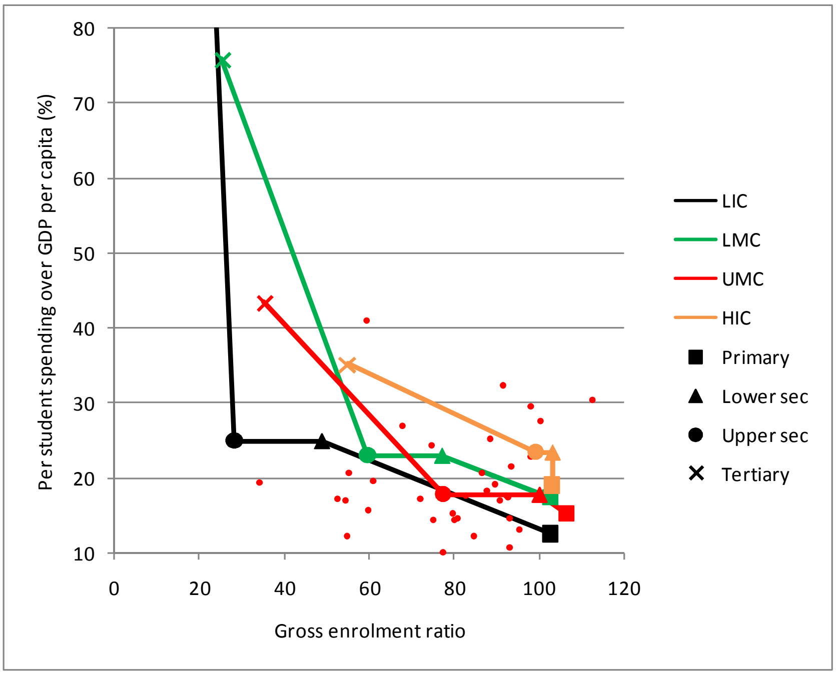
The patterns in the graphs seem unsurprising. On the expenditure side, the especially high level of resourcing devoted to each tertiary student, relative to GDP per capita, in low income countries stands out. This one might expect, given the globally competitive nature of tertiary education (developing country elites often send their young to rich country universities), meaning there is pressure to approximate the per student spending found in high income countries in absolute terms. To a lesser degree, this dynamic would influence the middle income tertiary spending ratios as well. On the enrolment side, if one examines the data points in the graphs against the horizontal axis, it is clear that as countries develop (keeping in mind, however, that these are cross-sectional data), the point at which large numbers of students leave the system shifts from primary schooling, in low income countries, to upper secondary schooling, in high income countries.

How much dispersion is there in the country values around the means behind the points of, say, Figure 38? The coefficients of variation in the next table indicate that there is more dispersion around the spending means than the enrolment means. Moreover, within each of the developing country categories, there is a particularly large degree of variation with respect to the tertiary spending indicator. A graphic illustration of the dispersion with respect to the two dimensions of one of the data points, namely that for upper secondary education in upper middle income countries, is given in Figure 39.

Table 27: Coefficients of variation for enrolment and spending statistics

	Gross enrolment ratio				Per student spending over GDP/capita			
Country category	Primary	Lower sec.	Upper sec.	Tertiary	Primary	Sec.	Tertiary	Countries
LIC	0.24	0.43	0.69	1.34	0.64	0.79	1.52	44
LMC	0.14	0.33	0.49	0.91	0.64	0.73	1.01	52
UMC	0.08	0.14	0.23	0.54	0.37	0.43	1.29	51
HIC	0.07	0.08	0.20	0.46	0.37	0.37	0.49	47
Overall	0.14	0.34	0.50	0.83	0.52	0.63	2.18	194

Figure 39: Figure 38 with dispersion illustrated

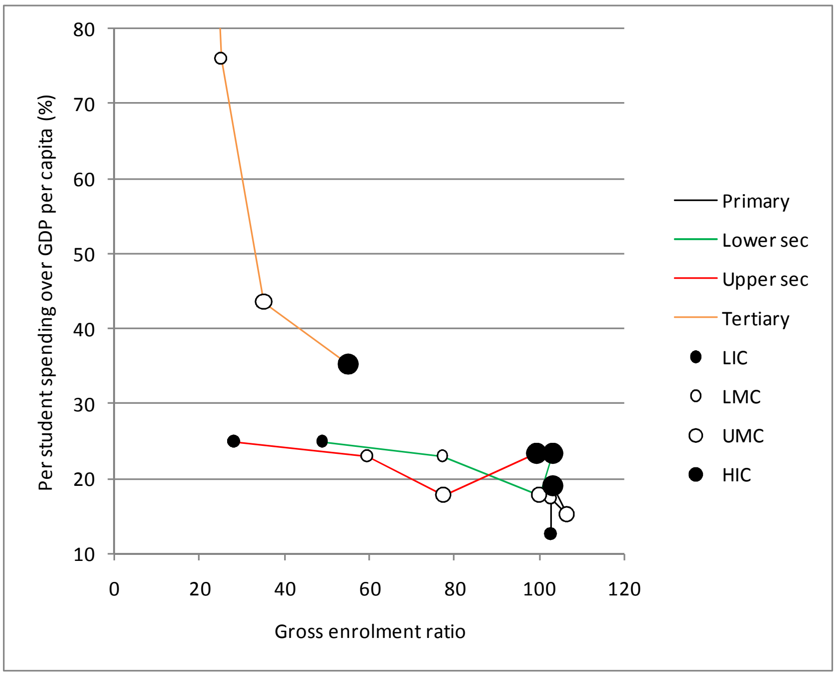


Note: Smaller points illustrate the dispersion of upper secondary values for upper middle income (UMC) countries.

To gain a better sense of the dispersion of country values with respect to the graphs appearing above, a calculation was run where it was determined whether each country-specific two dimensional point was closer to the correct mean point or some other mean point within a specific education level. For instance, if the upper secondary point for Zambia was closer, in a two-dimensional sense, to the lower middle income country mean point than the mean point for one of the other development categories, this Zambia point was given a score of 1 (Zambia is a lower middle income country). If not, it was given a score of zero. The mean across all the scores was 0.51, meaning that half of the time the country-specific points were closest to the mean point of their development category, the other half of the time they were not. What all this indicates is that the dispersion is substantial and that simplistic interpretations of the aggregate multi-country statistics, even where relatively robust patterns are found, should be avoided.

The previous three graphs, which are actually different views of the same graph, illustrate one of two key graph formats that will be used frequently in the analysis that follows. The second graph format is shown below and simply involves joining the points of the previous graphs differently, so that curves represent levels of education, where previously they represented country development categories. In Figure 40 below, we could say that curves point to the trajectory or pathway that, say, tertiary education can be expected to take as a country develops.

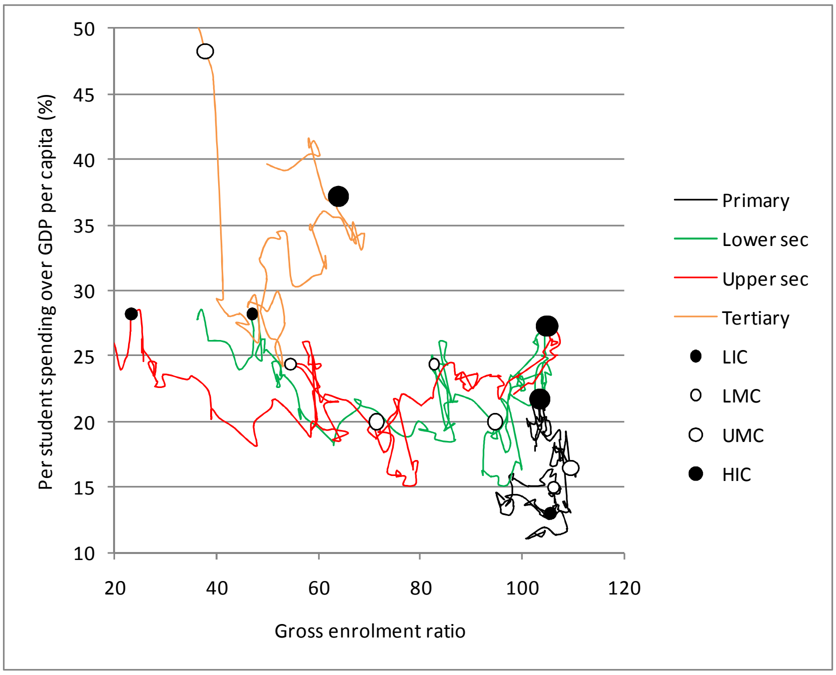
Figure 40: Figure 38 viewed transversally



Note: The points referring to development category (LIC etc.) indicate the means along the horizontal and vertical axes for the countries in the category in question.

It seemed worthwhile to explore a variation of the previous graph where more detail on the pathways would be seen, using not just the development categories of countries, but exact per capita income. The result was the following graph, admittedly not a very visually appealing graph, but it serves as a point of departure for a better graph that will be discussed below. Countries were placed in a list in ascending order of income, where this was GDP per capita in 2005 USD PPP terms (using the Penn data source discussed in section 10.1). As above, only data from a recent year not earlier than the year 2000 were used. This list was used to calculate moving averages with respect to the education indicators, where each average covered 21 countries. The lines in Figure 41 are lines joining the averages. Markers have been inserted at points along the pathways corresponding to countries with more or less the median income for a specific development category.

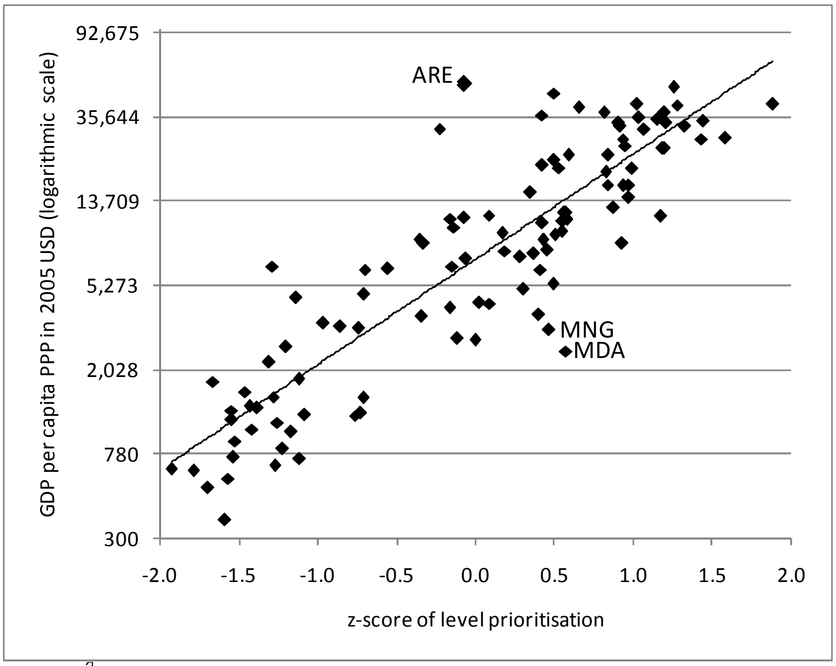
Figure 41: Moving average pathways for enrolment and spending



Note: Each marker is situated on a point reflecting the median per capita GDP value for each of the four development categories.

One problem with the foregoing graphs is that they do not consider patterns across all seven indicators within individual countries. Figure 41 does to some extent reflect the relationship across two indicators at a time within individual countries, for instance, primary enrolments and per student spending. A further problem with the last graph is that it assumes that income per capita is a reliable measure of a country's level of development. One way of dealing with these deficiencies is to completely ignore the income variable and the World Bank's development category and to seek a development trajectory using just the seven education variables. Factor analysis, using the factor command in Stata, allows us to do this. The 'principal factor' option for the command was used, and then using the post-estimation predict command, a single z-score was produced for each of the 112 countries included in the analysis, where this z-score in a sense summarised the seven indicator values. Only 112 countries could be included as each country had to have all seven values. The factor statistical outputs include uniqueness values for each of the seven input variables, which if below around 0.5 point to a high correlation between the variable and the z-score (StataCorp, 2011: 297). As one may expect from the above discussion, the primary GER is not a good predictor of a country's development status, but nor are the secondary and tertiary spending ratios. The other four variables, however, do display trends moving in the same general direction as the z-score. What is striking is that the z-score displays a high correlation with what would normally be considered a key indicator of country development, namely income per capita. The correlation coefficient if the untransformed income variable is used is 0.67, but it rises to 0.86 if the natural logarithm of income is used. The following graph illustrates the relationship.

Figure 42: Income and a measure of education development

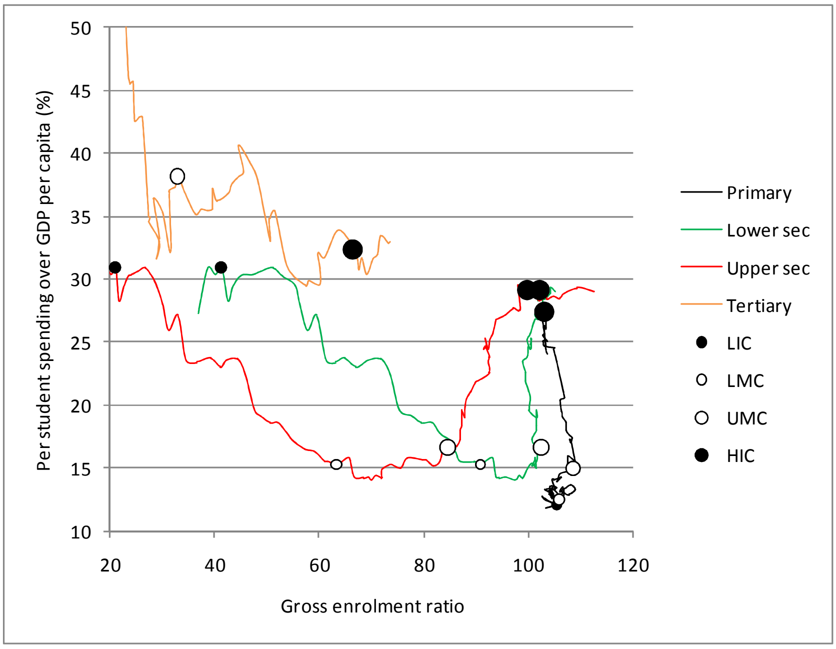


Note: R^2 for the trendline is 0.74.

Examining the outliers in the graph can be informative for policymakers. If one compares the seven values for United Arab Emirates (ARE) against those of countries with similar levels of income per capita, one finds that this country is an outlier because per student spending is low relative to GDP per capita. This fact, combined with other evidence, might point to spending on specific levels of education being too low. To the right of the trendline, Mongolia (MNG) and Moldova (MDA) are more developed according to their z-scores than their income largely because their level of tertiary enrolment is higher than that of other countries with similar incomes. The governments could use this fact to attract foreign investors, on the basis of their quality of human capital. Of course these kinds of policy conclusions can be drawn without the use of the graph, but the graph represents a novel and compact method of identifying whether countries are possible laggards or pioneers with respect to the global pattern, and with respect to the seven indicators in question.

Figure 41 can now be redrawn with countries ranked in the list according to a more educationally-focussed summary indicator of development, namely the z-score. The result is Figure 43 below. As one would expect, the curves now display fewer random movements because the 'noise' previously caused by the use of income as an indicator of development has been removed.

Figure 43: Moving average pathways without use of income

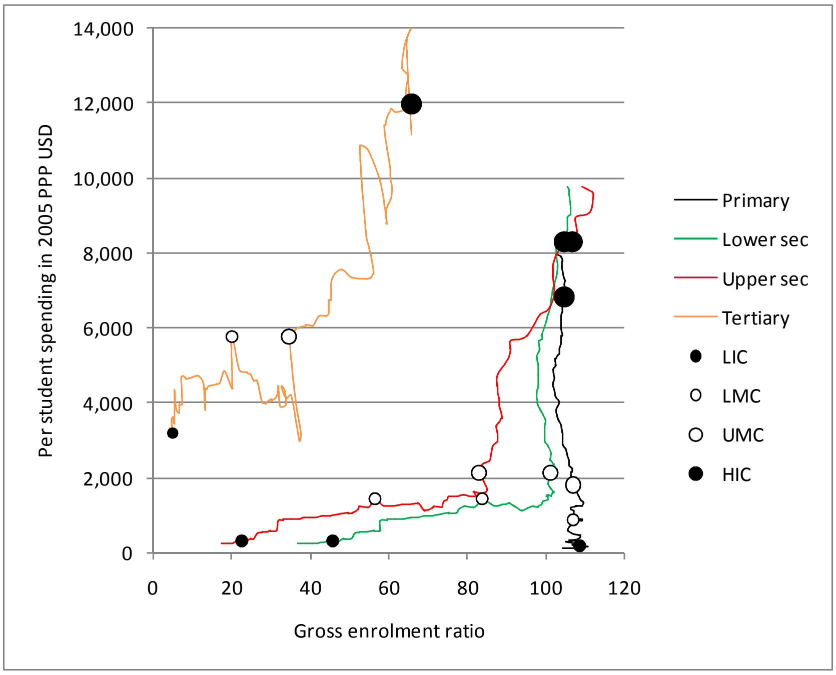


Note: Here, as in Figure 41, curves are smoothed using moving averages that cover 21 countries at a time. The points labelled 'LIC' etc. are positioned where they represent the median z-score per development category produced by the factor analysis.

One striking pattern seen above is that at the secondary levels there is a clear decline and then increase in spending per student relative to GDP per capita. Of course it should be kept in mind that this is a cross-sectional view of the data. When historical patterns are examined in the following section, different patterns are seen.

To conclude the analysis of cross-sectional patterns, the next graph takes the per student spending ratios from the previous graph and converts them into absolute monetary terms by multiplying them by GDP per capita. In absolute terms, there is a continuous increase in per student spending as countries develop, but the difference between the spending levels of upper middle income countries and high income countries is stark.

Figure 44: Figure 43 with absolute per student spending



10.4 Analysis of historical trends

Using cross-sectional data almost as if they revealed historical development trends, as was done in the previous section, is common. This is the approach employed by Mingat and Tan (2003), for instance. Of course historical trends should ideally be examined using historical data, and this is what is attempted in this section. One difficulty posed by a historical analysis is that even inflation-adjusted purchasing power parity per capita income values are not necessarily comparable over longer periods of time. This becomes clear if one examines, for instance, the income thresholds separating low income countries from middle income countries in the 1970s and in the 2000s. As will be seen below, they are very different. But before this complexity is discussed, the breakdown of countries according to development category appearing the World Bank's *World development report* for 1978 will be compared to that appearing in the 2012 report to bring to the fore complexities around inconsistent classifications, and inconsistent sets of countries (World Bank, 1978: 76; World Bank, 2011b).

The 1978 report of the World Bank used as its main categories 'industrialised countries', 'middle income countries' and 'low income countries', but these excluded 25 countries classified as 'centrally planned economies' and three classified as 'capital surplus oil exporters'. The following table shows the degree to which countries moved between the three main 1978 categories and the 2012 categories, where for the latter lower and upper middle income countries have been collapsed into one category, middle income country. The seven countries which moved from the middle income status in 1978 to the low income status in 2012 were six Sub-Saharan Africa countries plus Nicaragua. The five that moved in the opposite direction, from low to middle income status, were all South Asian or Southeast Asian countries. The one country that moved from industrialised country to middle income country status was South Africa, but an examination of income values reveals that in 1978 South

Africa's per capita income was well below that of any other industrialised country, suggesting that anomalous criteria were used when South Africa was classified. All other industrialised countries in 1978 can clearly be considered high income countries too. Finally, the ten countries that moved from the middle to high income status included five Southern European countries, four East Asian countries and Israel.

Table 28: 1978-2012 country classification changes I

2012 ➔ 1978 ➔	LIC	MIC	HIC	Totals 1978
LIC	28	5		33
MIC	7	44	10	61
IC		1	18	19
Totals 2012	35	50	28	113

Note: In this and the following table, 'IC' refers to 'industrialised countries'.

For the historical analysis it seemed important to impute, where possible, 1970s country development categories according to the three main income groups, where such categorisations had not been made for 1978. The first step was to take the 1978 categories of countries that ceased to exist, in particular the Soviet Union, and to assign the same categories to the 'splinter country', for instance Ukraine. The next step was to identify income levels that could be used to separate low from middle, and middle from high income, in 1970, the earliest year for which the education data were available, but also 2009, the last year for which the Penn income data were available, using the 1978 and 2012 categories respectively. It was thus assumed, for instance, that categories would not have changed between 1970 and 1978. The Penn income values from 1970 and 2009 were taken and points found at which the least number of countries would find themselves on the wrong side of the divide. For instance, in establishing an income value to divide low income countries from middle income countries in the 1970 data, a value was found that minimised the number of low income countries with an income above that level plus the number of middle income countries with an income below that level. It was found that the income level dividing low from middle income countries was 1,200 USD in 1970 and 2,300 USD in 2009. This is despite the fact that income values were real inflation-adjusted PPP values, which might create the impression that thresholds between country development categories should remain static over time. That this should not be the case is unsurprising if one considers the great conceptual and technical difficulties that exist in making prices comparable over time and space. This was discussed in Part I of the dissertation.

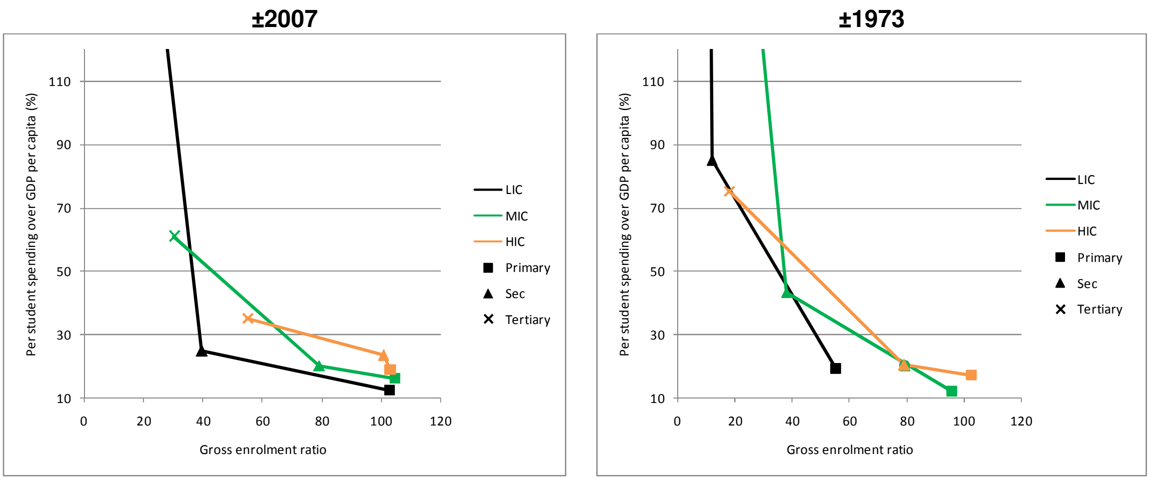
The division between the middle and high income categories was found to be 10,000 USD for 1970 and 24,000 USD for 2009. Here industrialised country was considered as synonymous with high income country, but South Africa was considered a middle income country in the 1970s. The 1970 income thresholds were used to place the 7 'centrally planned economies', or their splinter countries, which had Penn income data for 1970, into one of the three income categories. Then the remaining centrally planned economies and 'capital surplus oil exporters' were categorised using the earliest available income value (the mean year was 1991) and the 2009 income threshold values. This method was not ideal, but the resultant 1970s classifications seemed appropriate. Finally, 43 countries, mainly from Africa, which did not appear in the World Bank's 1978 list but which did have 1970 Penn income values were categorised, using the 1970 income thresholds. The result, shown in Table 29 below, was 184 countries which had categories for both 1978 and 2012.

Table 29: 1978-2012 country classification changes II

2012 ➔ 1978 ➔	LIC	MIC	HIC	Totals 1978
LIC	34	13		47
MIC	9	86	14	109
IC		2	26	28
Totals 2012	43	101	40	184

The left-hand graph below is a reproduction of Figure 37. The right-hand graph represents the comparable situation using the oldest UIS data available and the country categorisations for the 1970s, derived as explained above. Along the lines of the approach used for the 2007 graph, for the right-hand graph a score of data availability per country and per year was calculated and the earliest year with the maximum score, as long as the year was before 1980, was used for each country. The mean year used across countries was 1973.0. The 1973 graph reflects 168 countries, against 194 for 2007. If one were to use the assumption, clearly a problematic one, that income values from 1970 are comparable to those of 2009, and that one can thus use the same thresholds between income categories in 1973 as one does in 2007, the resultant 1973 would look different, but it would still be closer to the right-hand graph below than the left-hand graph. The conclusion we can thus draw is that however one views the data, there have been large shifts in recent decades in the way similarly developed countries prioritise education levels.

Figure 45: Education levels in 2007 and 1973



The following table presents the values behind the above graphs. The general trend is that the enrolment values have risen and that expenditure values have dropped. On the enrolment side the move towards universal primary schooling is not surprising, given the prevailing policy pressures, but that secondary and tertiary GERs should have increased to the extent they have is remarkable. For instance, enrolment at the tertiary level more than tripled in all three development categories. Enrolments at the secondary level in low and middle income countries more than doubled. As seen in the last three rows of the table, the reductions in the spending indicator values are proportionally strikingly similar to the increases in the GERs, suggesting that there has been a trade-off between high spending and higher enrolments. If one ignores changes in the distribution of the population over ages, the closer the ratios in the bottom left panel of the table are to the corresponding ones in the bottom right panel, the stronger the likelihood that the overall level of investment in education relative to GDP remained the same, and that similar proportions of GDP were simply spread out over more students. As will be discussed in the conclusion in section 11, historical trends of the kind displayed here have received considerable attention in the economic literature, yet how optimal they have been remains an unresolved matter.

Table 30: Education levels in 2007 and 1973

	Gross enrolment ratio			Per student exp. / GDP per capita		
	Primary	Sec.	Tertiary	Primary	Sec.	Tertiary
2007						
LIC	103.0	39.5	9.2	12.5	24.8	276.9
MIC	104.5	79.1	30.1	16.2	20.1	61.1
HIC	103.0	100.6	54.8	19.0	23.5	35.3
1973						
LIC	55.2	12.0	1.4	19.3	85.0	1314.3
MIC	95.8	38.0	8.1	12.1	43.4	322.1
HIC	102.5	79.0	17.7	17.3	20.4	75.6
Change	**2007 over 1973**			**1973 over 2007**		
LIC	1.9	3.3	6.8	1.5	3.4	4.7
MIC	1.1	2.1	3.7	0.7	2.2	5.3
HIC	1.0	1.3	3.1	0.9	0.9	2.1

One complication if one wants to analyse the income and education indicator data described in section 10.1 as a panel dataset is that the meaning of the income values change over time, as mentioned previously. The same supposedly real income value can mean a middle income in 1970, but a low income in 2009, for example. The complication is not insignificant if one considers that the shifting thresholds referred to above imply roughly a halving the value of the supposedly constant USD over four decades, if one assumes that, for instance, the difference between a low income country and a middle income country means in the 2000s what it did in the 1970s. An attempt was made to use the income thresholds separating the development categories discussed earlier to adjust the Penn income data so that, say, 10,000 USD would mean the same thing in 1970 and 2009, in terms of the its location within the country development categories. However, this attempt was abandoned as the adjusted values obtained introduced counter-intuitive results, in particular negative GDP per capita growth for many countries and many years where the original values reflected positive growth.

A more serious and fundamental complication if one wants analyse the historical patterns in a manner that goes beyond a simply descriptive approach, as seen in Table 30, is that the historical period reflected in the data is too short to provide a robust picture of what typically occurs to the education indicators of a country over a longer development trajectory. This is not just a data problem. Country development as one might understand it in development economics has not been occurring long enough for very neat patterns to be extractable from the historical data. It is of course debatable whether changing technologies and the more recent phenomenon of a changing ecology would ever allow neat and consistent patterns of country development to emerge. This is why development economics has so often turned to cross-sectional analysis in attempting to identify the development trajectories that appear to prevail at a specific historical point. Yet the short historical series we have in the data present some opportunities, in particular the opportunity to examine whether the cross-sectional patterns are replicated in the longitudinal patterns.

In the data we have, the cross-sectional patterns would tend to dominate over the historical patterns if, for instance, one simply regressed the secondary GER on income. The coefficient on income would describe differences across countries to a much larger degree than differences within countries across years. The intraclass correlation coefficient if one takes all the historical values and treats country as one's class is high, for instance 0.817 for the secondary GER and 0.621 for the tertiary GER. Thus 82% of the variation of one's secondary GER values is explained by the country, as opposed to the year. The econometrics answer to the need to separate cross-sectional from historical trends is a fixed effects model. The fixed effects model represented by equation (88) below was used. W is the indicator value for country i and year t. Apart from the overall intercept λ_0, each of the n countries except for one also has an intercept S. Thereafter there are three slope coefficients β linked to the natural logarithm of income per capita I, growth g in I between t and t minus 1, and the interaction between I and g. The results are presented in Table 31. For each of the 11 indicators, one of four variants of equation (88) was used. Both the untransformed version of the dependent variable and its natural logarithm were tested. Models were moreover run with and without the interaction term. The model with the interaction term was rejected if this term's coefficient was not significant at the 10% level. Thereafter the model with the highest R^2 value was used. One thing that should be noted about the results is that annual growth in income displays a statistically significant and independent association with all the indicator values. This points to different trajectories for the indicator values for faster and slower growing economies. The shape and magnitude of this association will be explored further in the next section.

$$W_{it} = \hat{\lambda}_0 + \left(\hat{\lambda}_2 S_{i=2,} + \ldots + \hat{\lambda}_n S_{i=n}\right) + \hat{\beta}_1 \ln I_{it} + \hat{\beta}_2 g_{it} + \hat{\beta}_1 \ln I_{it} g_{it} + \hat{u}_{it} \quad (88)$$

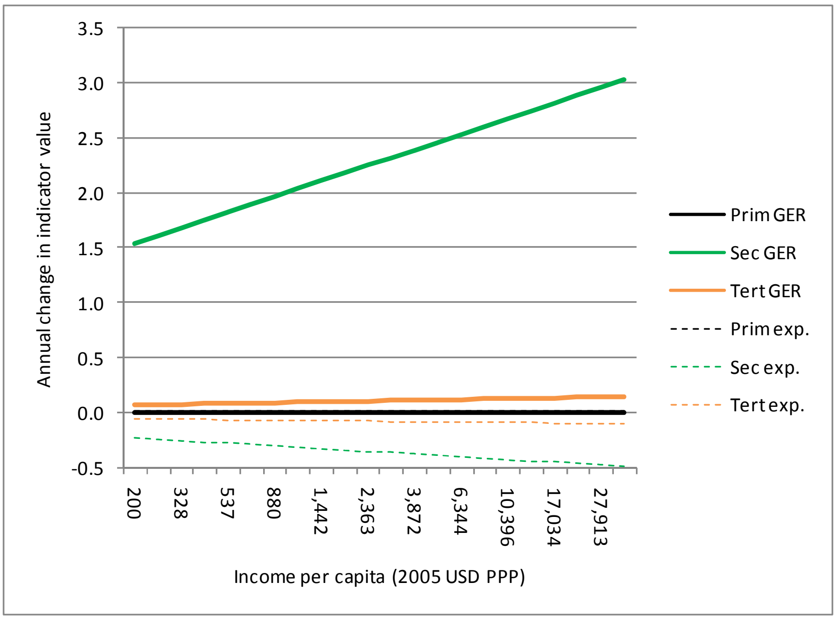
Table 31: Regression of indicator values on historical income

			Coefficient on...			
Variable	Obser-vations	Cons-tant	Nat. log of income	Growth	Interaction of prev. two	R^2
GER primary (ln)	5562	3.9	0.1***	1.4***	-0.2***	0.02
GER lower secondary (ln)	1780	1.0	0.4***	1.3***	-0.2***	0.19
GER upper secondary (ln)	1697	0.2	0.4***	1.5***	-0.2***	0.15
GER secondary	4832	-178.1	28.1***	37.8**	-7.2***	0.34
GER tertiary (ln)	4008	-9.3	1.4***	2.5***	-0.4***	0.40
Vocational ISCED 2/3	4193	27.2	-1.4***	-2.6*		0.01
Post-secondary non-tert. (ln)	750	12.4	-1.2***	1.2***		0.13
Tertiary engineering	733	61.8	-5.2***	-3.4*		0.10
Per student exp. primary (ln)	2087	0.7	0.2***	-0.3**		0.04
Per student exp. sec.	2021	66.7	-4.4***	89.1*	-10.2*	0.01
Per student exp. tertiary (ln)	2013	12.4	-0.9***	-0.6***		0.24

*Note: '(ln)' refers to the fact that the dependent variable was expressed as a natural logarithm. The per student expenditure values expressed as a ratio to GDP per capita. Regression results are from a model with fixed effects for countries. *** indicates that the estimate is significant at the 1% level of significance, ** at the 5% level, and * at the 10% level. Growth in the model was expressed as a fraction, e.g. 0.03. R^2 values refer to within-country explanatory power. The overall R^2 would be higher.*

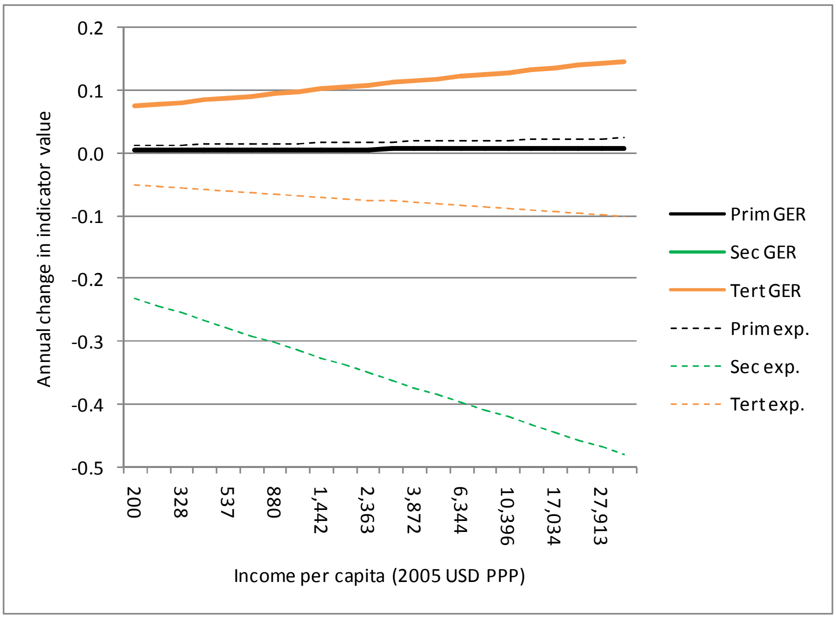
The following two graphs illustrate the average annual gains or losses for the indicators dealing with the three key education levels, at different income levels, using the slope coefficients from the previous table. An annual growth in income per capita of 2.5% was assumed. The patterns seen in the graphs are broadly in agreement with the cross-sectional findings discussed in section 10.3 in the sense that secondary and tertiary enrolment GERs increase with development and that the spending ratios decrease. One anomaly, however, is that the conclusion one may draw from the cross-sectional picture (Figure 43, for instance) that the primary and secondary spending ratios increase at high levels of development, is not supported by the historical trends in the graphs appearing below, and nor is it suggested by the non-parametric analysis presented below. Moreover, the large relative magnitudes of the secondary enrolment and spending shifts seen below are not supported by the cross-sectional picture. For instance, if one uses just the 2007 values in Table 30 and assumes an annual growth in income of 2.5%, between the LIC and LMC medians one might expect the annual increase in the secondary GER to be 0.8, against the 2.0 suggested by the historical trend in Figure 46.

Figure 46: Historical trends for education indicator (parametric)



Note: Income values on the horizontal axis are expressed on an exponential scale based on a 2.5% annual growth rate.

Figure 47: Magnification of Figure 46



The 'Vocational ISCED 2/3' indicator, for which we have values covering four decades (Table 25), comes with a negative coefficient on income in Table 31. However, this indicator refers to the percentage of secondary level enrolments in vocational streams. If one multiplies it by the secondary GER and then re-runs the model, the coefficient on income is positive. Vocational enrolments at the secondary level thus tend to increase relative to population as a country develops, but not as fast as non-vocational secondary enrolments. To illustrate the general trend, across all countries the mean percentage of secondary enrolments which are vocational dropped from 16% to 14% between the 1970s and the 2000s, but secondary vocational enrolments relative to the population increased from 6% to 12% over the same period.

A similar pattern is seen for the post-secondary non-tertiary and tertiary engineering indicators, for which data are available only since 1998. For the former, the shift between the 1998 and 2012 was from 15% to 13%, whilst as a percentage of the population enrolments moved from 3.5% to 4.0%. Tertiary engineering enrolments as a percentage of the age-relevant population moved from 5.7% to 6.5% over the same period.

One limitation of the fixed effects models discussed above is that the coefficients on income can be said to capture both a time invariant income correlation and the historical effect discussed earlier whereby, for instance, large increases in tertiary enrolments occurred in developed countries beyond what seems attributable to income. The models reflected in Table 31 were rerun with three decade dummy variables inserted: whether a year was 1980 or later, 1990 or later, or 2000 or later. These dummies might provide an indication of the magnitude and timing of historical effects. Obviously the correlation between the dummy variables and income within each country would be high and thus multicollinearity in the explanatory variables would be high. This should caution one against reading too much into a comparison between the coefficients on the income variable and those on the dummies seen in Table 32. Yet it seems noteworthy that in general the coefficients in all columns are simultaneously significant, across all indicators. What is perhaps most telling about the results is when the large historical trends occurred. Primary enrolments received a boost, beyond what income levels would predict, in the 1980s, and again after 2000. For secondary enrolments, both the 1980s and 1990s saw large upward movements. Tertiary enrolments increased above all in the 1990s. In general, the downward trends in the expenditure ratios occurred when the upward trends in the GERs occurred, making it plausible that enrolments could expand partly because trends in spending per student were contained, or did not increase as fast as GDP per capita.

Table 32: Regression of indicator values on income and decades

	Coefficient on...						
Variable	Nat. log of income	Growth	Interaction of prev. two	Dummy for since 1980	Dummy for since 1990	Dummy for since 2000	R^2
GER primary (ln)	-0.1***	69.7***	-7.4***	7.3***	1.4***	6.9***	0.164
GER lower secondary (ln)	17.0***	29.4	-4.2*			4.0***	0.213
GER upper secondary (ln)	10.6***	1.9	-0.2			0.1***	0.178
GER secondary	11.9***	11.9**	-2.9*	8.8***	8.3***	0.2***	0.586
GER tertiary (ln)	13.0***	-7.5	-0.2	3.0***	5.5***	0.4***	0.680
Vocational ISCED 2/3	0.2***	16.9*	-2.5**	-0.1***	-1.0***	-0.2***	0.043
Post-secondary non-tert. (ln)	-1.1***	70.9***	-6.7**			-0.1	0.128
Tertiary engineering	-0.2***	-9.5	0.7			0.0	0.105
Per student exp. primary (ln)	1.9***	21.8	-3.3**	-0.1	0.1	0.1***	0.045
Per student exp. sec.	0.2***	114.3	-14.3***	-10.6***	-3.7*	0.0	0.087
Per student exp. tertiary (ln)	-0.3***	941.5	-127.9	-0.3***	-84.7***	-0.2***	0.361

*Note: '(ln)' refers to the fact that the dependent variable was expressed as a natural logarithm. The per student expenditure values expressed as a ratio to GDP per capita. Regression results are from a model with fixed effects for countries. *** indicates that the estimate is significant at the 1% level of significance, ** at the 5% level, and * at the 10% level. Growth in the model was expressed as a fraction, e.g. 0.03. R^2 values refer to within-country explanatory power. The overall R^2 would be higher.*

What might be more useful for policymakers than the parametric patterns described above would be a non-parametric, or semi-parametric, graphing of the indicator trends over time, relative to income, to reflect irregular variation. A convenient tool for this are curves with lowess smoothing. The next two graphs use such curves, with the 'bandwidth' set at the Stata default of 0.8, meaning that the locally weighted regression analysis run for each point on the curve uses 80% of all observations, with more distant observations, with respect to the x-axis, being weighted much less[4]. The *x* values are the Penn USD income values at 2005 PPP prices. The *y* values were obtained by running multiple regressions in the dataset. In each regression, data from just five consecutive years within one country were used, with the indicator values being regressed on year to obtain the slope. The resultant coefficient on year would then be attached to the last year. So, for instance, the 2006 secondary GER slope for South Africa would be the slope of the indicator values on year for 2002 to 2006.

[4] Of course there must be enough observations on either side of the value within the sorted independent variable to fulfil the 0.8 condition. In fact, below the 40th percentile of the independent variable and above the 60th percentile, fewer than 80% of observations will be used in the local regression.

Figure 48: Lowess curves for historical indicator trends against income

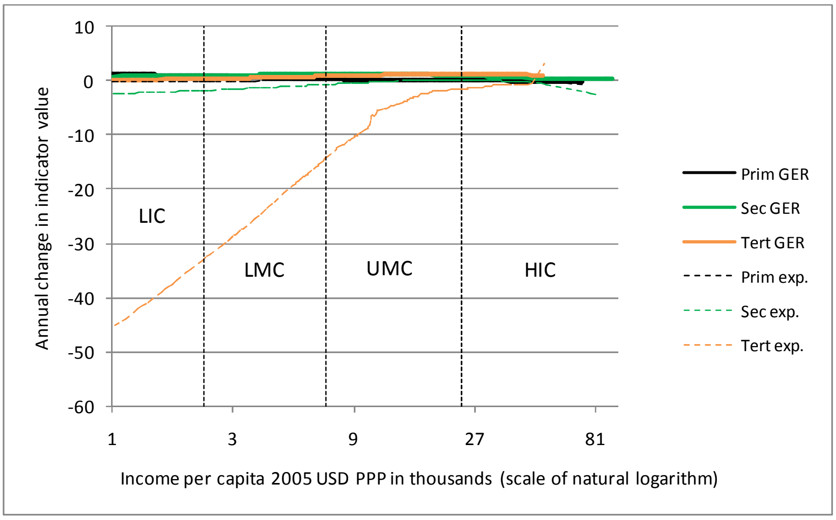
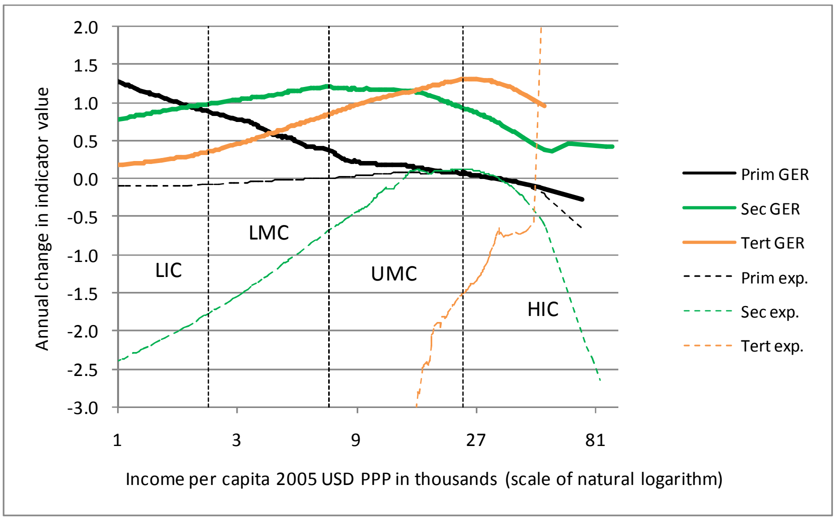


Figure 49: Magnification of Figure 48



The pattern seen earlier of decreasing secondary and tertiary expenditure ratios as income rises appears again, except here these decreases are reduced over time whilst in Figure 47 they appeared to become larger over time. Moreover, decreases in the primary GER as income increases, presumably as inefficiencies in the form of high grade repetition are resolved, become clear in the above graph, whilst this did not emerge from the parametric analysis.

If we redo the last two graphs using only the *y* values for 2000 and later, in order to provide a picture of what can be considered current trends, we obtain the following two graphs. Using only the recent data results in larger secondary and tertiary GER increases and smaller reductions in the expenditure ratios for these education levels.

Figure 50: Lowess curves for just 2000 and beyond

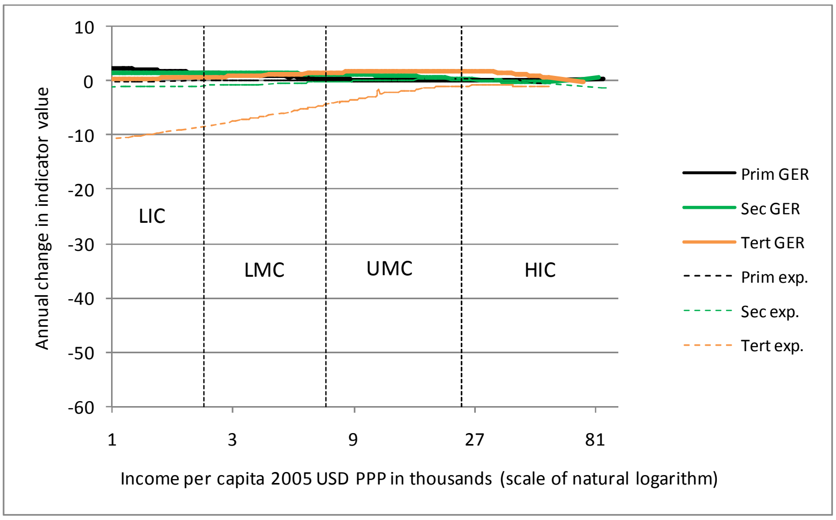
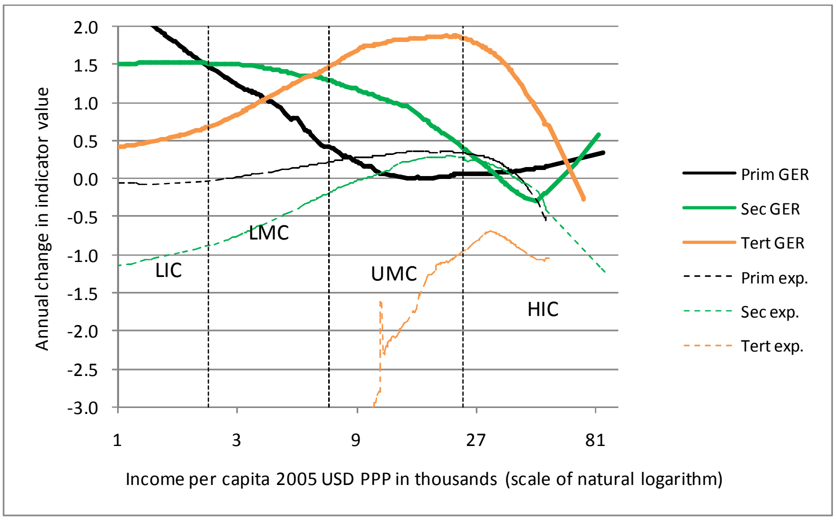


Figure 51: Magnification of Figure 50

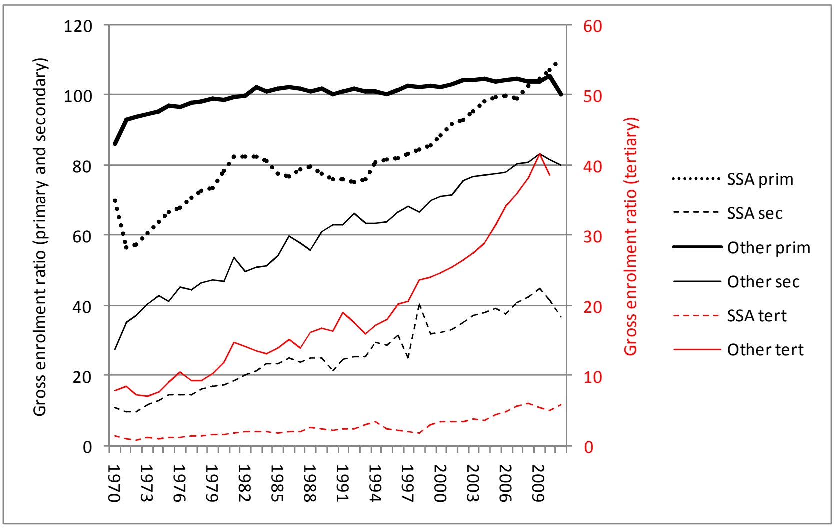


What has been presented above is important in terms of uncovering basic patterns in the data. A more policy-oriented discussion using the graphing methods developed here appears in the next section. One region-specific policy matter concludes this section, namely the question of whether the past prioritisation of primary schooling by international agencies in the case of Sub-Saharan Africa (SSA) led to the under-prioritisation of the secondary and tertiary levels. Some examination of the GER values suggests that one's answer to this policy question could be yes or no, depending on how one interprets the data. The following two graphs use GER values for countries that are not high income countries. Figure 52 uses all the values available, whilst Figure 53 uses the values of just those countries where at least 80% of the values were non-missing[5]. The similarity of the two graphs is remarkable if one considers that only a quarter of the number of SSA countries is used for the second

[5] Specifically, 126 possible values per country were considered, as there were three indicators and 42 years. If no more than 25 values of the 126 were missing, the country was included in the calculation (25 being 20% of 126).

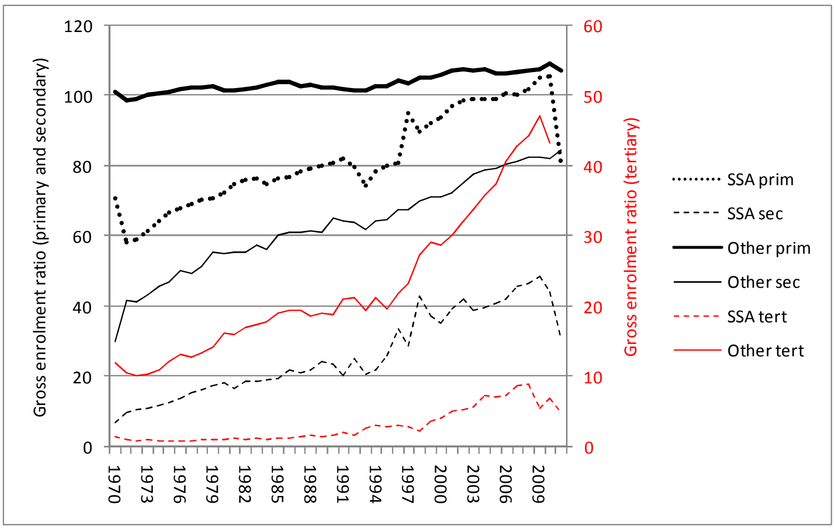
graph compared to the first graph. The graphs use data from up to around four Sub-Saharan African upper middle income countries, and around half of the countries from the other regions are upper middle income countries. However, this fact does not influence the overall picture. If upper middle income countries are excluded from the graphs, leaving just low and lower middle income countries, the picture remains virtually unchanged.

Figure 52: Sub-Saharan Africa's prioritisation of education levels I



Note: The average number of countries per yearly data point for SSA is 38, 30 and 21 for the primary, secondary and tertiary levels. The figures for non-SSA are 72, 64 and 54.

Figure 53: Sub-Saharan Africa's prioritisation of education levels II



Note: The average number of countries per yearly data point for SSA is 8, 7 and 6 for the primary, secondary and tertiary levels. The figures for non-SSA are 25, 24 and 23.

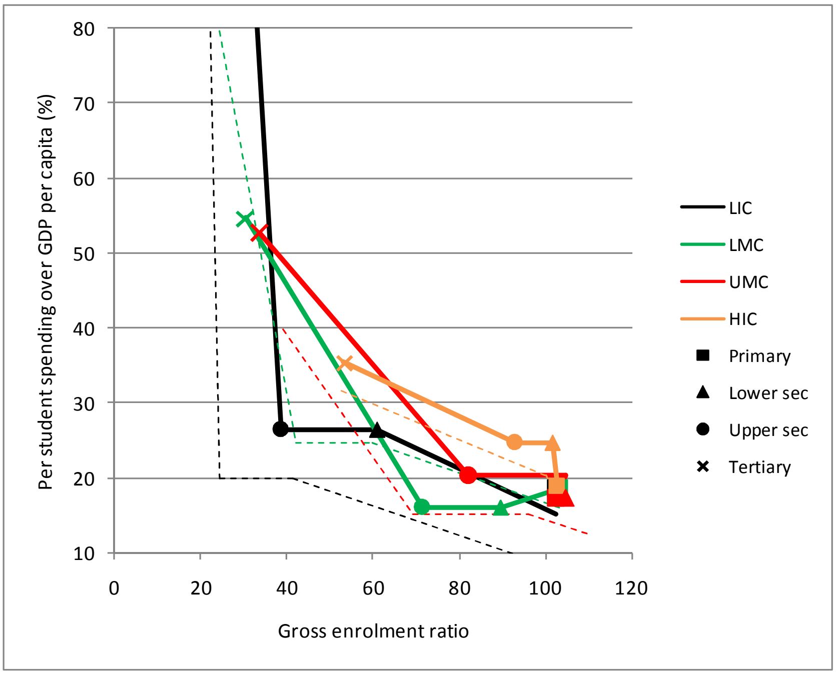
In defence of the actual historical trend, one could argue that the SSA region succeeded in bringing its primary GER up to the level of other developing regions over the four decades, whilst tertiary enrolments also increased substantially, albeit off a low base. Specifically, tertiary GER values increased by a factor of 6.5 over the four decades in the region against 4.5 for other countries (ratios taken from the second graph). Similarly, the increase in secondary enrolments was steeper in SSA than elsewhere. The ratio of the primary GER to the secondary GER declined from 59 to 14, whilst the ratio of the secondary GER to the tertiary GER declined from 6 to 2. These ratios shifted to a greater degree in the region than elsewhere. The proportion of SSA students enrolled at the primary level, as opposed to some other level, was thus far greater in 1970 than in 2011. But one could also use the data to support the opposite argument. By 2011, whilst the SSA region was no longer behind with respect to primary enrolments, its secondary enrolments were at around half of the level found in comparator countries, whilst it lagged even further behind with respect to tertiary enrolments. One may ask whether the priority should not rather have been to close the gaps in a manner that focussed less on universal primary schooling, and more on balancing enrolments across levels, using international patterns as a norm, and thus for instance accepting that universal primary schooling would not be reached whilst secondary enrolment levels were still so low. The last two years in the first graph, 2010 and 2011, display a situation where the primary GER in SSA was slightly *higher* than that elsewhere. If this trend continues, the argument that the region has over-prioritised primary schooling could be strengthened.

10.5 Growth and education prioritisation patterns

Sections 10.3 and 10.4 have explored ways of viewing and analysing the UIS data in combination with the Penn income data. They also served as an important check of anomalies in the data, for instance the changing meaning of the real income values over time. On the whole it was found that the UIS data, despite gaps, are sufficiently complete and correct to produce meaningful patterns that could be of interest to policymakers. Below, how the patterns seen in the UIS data differ by level of economic performance is examined. The implications for policymakers are also discussed. Economic performance is mostly defined in terms of increases in the Penn income per capita values. However, where indicators are inserted into the semi-Bayesian model of Sala-i-Martin, Doppelhofer and Miller (2004), performance is gauged in terms of overall GDP growth.

The following graph follows the format of Figure 38, but provides separate curves for the top and bottom terciles of performers. Performance per country is the average annual percentage increase in income per capita, found by regressing the natural logarithm of the Penn income values on year. The top and bottom terciles were identified within each of the four development categories applicable in 2012. The data behind the next graph are presented in Table 33. The table is perhaps most convenient for interpreting the trends. The last four rows of this table show the difference with respect to the indicator values between the faster and slower growing countries after income has been controlled for. Specifically, within each development category, and with each observation being one country, the indicator value was regressed on the natural logarithm of 2007 income and two dummies, one indicating that the country was a fast grower, the second that the country was a slow grower. Each difference value at the end of the table is the coefficient on the fast growth dummy minus the coefficient on the slow growth dummy.

Figure 54: Educational level prioritisation ±2007 by growth category



Note: This graph is a modification of Figure 38. Solid lines represent better performing counties, whilst dotted lines represent worse performing countries.

Table 33: Indicator values ±2007 by growth category

	Gross enrolment ratio				Per student spending over GDP/capita			
Country category	Primary	Lower sec.	Upper sec.	Tertiary	Primary	Sec.	Tertiary	Countries
Within top tercile with respect to 1970-2009 growth								
LIC	103	61	39	15	15	26	254	14
LMC	103	90	72	30	19	16	55	18
UMC	103	105	82	34	17	20	53	18
HIC	103	102	93	53	19	25	35	15
Within bottom tercile with respect to 1970-2009 growth								
LIC	95	41	24	6	9	20	534	13
LMC	103	60	42	12	16	25	117	16
UMC	110	96	69	39	13	15	40	16
HIC	103	100	100	53	16	20	32	14
Difference between top and bottom terciles (positive means value in best tercile is higher)								
LIC	7	20	14	9	6	6	-280	
LMC	0	30	30	18	2	-8	-62	
UMC	-7	9	13	-5	5	5	13	
HIC	-1	2	-7	0	3	4	4	
Difference between top and bottom terciles after controlling for income								
LIC	9	11	7	2	10	14	-189	
LMC	-4	18	15	4	-3	-8	-24	
UMC	-6	8	9	-7	3	3	16	
HIC	-1	2	-7	-2	3	5	7	

The general patterns hold whether one controls for income or not, though differences are, as one might expect, reduced when the income control is introduced. What are these general patterns? Better performing countries tend to have tertiary education indicator values, on both the enrolment and spending sides. These countries thus prioritise education more. The pattern becomes more pronounced the lower the development category. There is one notable exception. Amongst lower income countries, those that grow slowly tend to have a higher tertiary education spending ratio. The fact that the general pattern of higher spending ratios, at all education levels, in faster growing economies should apply in high income countries, seems noteworthy. Work such as that by Krueger and Lindahl (2001) suggests that high income countries stand to gain little from investing more in education. Of course the statistics in Table 33 say nothing conclusive about cause and effect. But one plausible interpretation is that well-governed countries, in terms of growth, are also better at managing their education systems, and understand that certain spending thresholds must be met if the education system is to support development optimally.

The fact that low and lower middle income countries that grow slowly display higher spending ratios for tertiary education warrants some closer scrutiny. One hypothesis would be that these countries are smaller countries where costs are inflated because economies of scale cannot be attained. The following correlation matrix, which ensures that for every pair of variables all observations with non-missing values in the pair are included, provides some initial answers. As already seen, lower growth is associated with higher spending ratios at the tertiary level. Moreover, lower growth is indeed associated with smaller populations and population is negatively correlated with the tertiary spending ratio. But correlations are extremely low. Moreover, smaller countries also have higher primary and secondary spending ratios and here one would not expect an economy of scale problem.

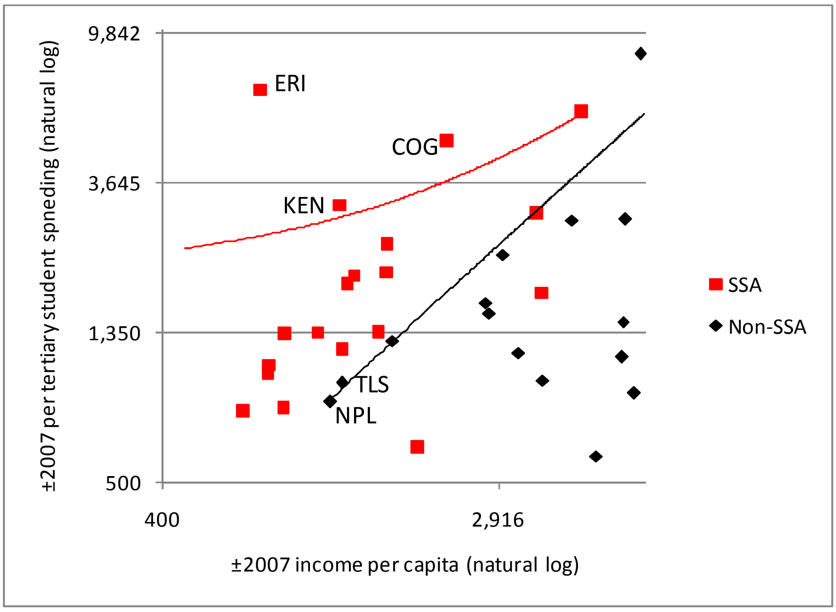
Table 34: Correlation matrix for LIC and LMC education spending

	Natural log of income	Average annual growth	Population	Primary spending	Secondary spending	Tertiary spending
Natural log of income	1.00					
Average annual growth	0.58	1.00				
Population	0.04	0.11	1.00			
Primary spending	0.34	0.14	-0.10	1.00		
Secondary spending	-0.10	-0.04	-0.10	0.26	1.00	
Tertiary spending	-0.33	-0.23	-0.08	-0.06	0.29	1.00

Note: Education indicator and income values used were those of around 2007. Growth is for the period 1970-2009.
Source: As for previous tables, but with population being 2010 estimates drawn from the online data-querying facility for the 2008 World Population Prospects of the UNSD.

What is revealing is to examine the regional dynamics behind the high tertiary spending phenomenon. The natural logarithm of USD tertiary spending per student, where the USD value was found by multiplying the tertiary spending ratio by GDP per capita, was regressed on the natural log of income, total population and dummies for world regions used by the World Bank. Only countries with an income value of 7,000 or below was included in the model. The coefficient on population was found to be insignificant. However, the coefficient on the Sub-Saharan Africa (SSA) dummy was positive and significant, regardless of the excluded dummy, whilst Eastern Europe, Latin America and East Asia and Pacific emerged as having low tertiary spending values. Even if all the regional dummies were excluded, population was still not a statistically significant predictor. The phenomenon of high tertiary unit costs amongst poorer nations seen in the data seems to be driven largely by what happens in the SSA region. This can be seen in Figure 55 below. The patterns point to a need within many SSA countries to restructure the way tertiary education is offered. There could be an artificial economy of scale problem in these countries in the sense that enrolments at the tertiary level have been low, partly for structural reasons which include low secondary enrolments (see Figure 52), which in turn could have contributed towards uneconomically small universities.

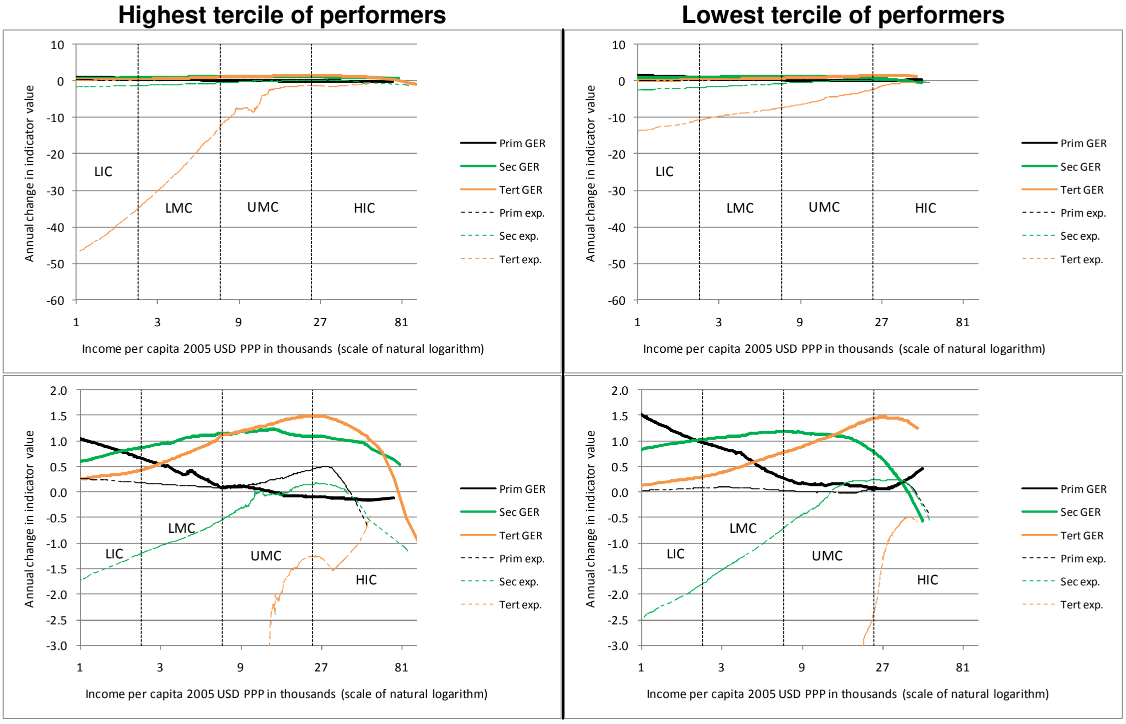
Figure 55: The SSA region and spending on tertiary education



Note: Trendlines are only for countries with less than a per capita income of USD 7,000 in 2005 PPP terms in ±2007. The graph magnifies the bottom left part of the graph and does not display all pairs of data points. Country codes are ISO codes – see Appendix A of Part I of the dissertation for explanations.

The graphs reflecting historical trends in the data discussed in the previous section were reproduced for better and worse economic performers (using the performance criteria explained earlier). The following graphs (which are comparable to Figure 48 and Figure 49 above) are interesting mostly with respect to secondary and tertiary spending trends. At the low and lower middle income levels, it is the faster growing economies that have seen the largest annual reductions in their tertiary spending ratios. This would agree with the pattern seen earlier of on average lower tertiary spending ratios for less developed countries in the 2000s that had grown faster. The secondary spending ratio, in contrast, has declined faster in economies that grew more slowly, at least at the low and lower middle income levels. Table 33 revealed a mixed picture for the 2000s, with low income countries displaying a positive correlation between growth and the secondary spending ratio, whilst the correlation in lower middle income countries was negative.

Figure 56: Lowess curves for historical trends by economic performance



A rather good way of testing the conditional correlation between the education indicators discussed in the previous sections and economic growth is to insert these indicators into the Bayesian averaging of classical estimates (BACE) model of Sala-i-Martin, Doppelhofer and Miller (2004), henceforth referred to as SDM. This technique was discussed in detail in Part I of the dissertation. If early values for the indicators are prioritised by the model, this could have policy implications and provide guidance for the kinds of explanatory variables to select in cross-country growth regressions.

In proceeding with the BACE analysis, it was necessary to consider the trade-offs between keeping education variables and keeping countries in the dataset to be used, given that the model cannot handle missing values. Moreover, various permutations of the indicator values were tested. In the end, after some preliminary testing of various education variables in the BACE process, it was decided to insert five variables into the original SDM model: the average annual change in the secondary and tertiary spending ratios (over the four decades), and the natural logarithm of the absolute USD value of per student spending in the 1970s for the primary, secondary and tertiary levels. The last three variables were obtained by multiplying the UIS spending ratio by GDP per capita, and the mean across the earliest three values from the 1970s was used. Enrolment ratios did generally not perform well in the SDM model, and clearly not as well as the derived spending variables. The 1970s primary GER did perform relatively well, but not as well as the 1960 primary GER used in the original SDM work.

Results for explanatory variables displaying high probabilities of being included in an optimal model, according to the BACE methodology, are displayed in Table 35 below[6].

Table 35: BACE regression outputs – growth on UIS and other data

	New calculation		Original SDM	
Var.	PIP	Coeff.	PIP	Coeff.
East Asian dummy	0.970	2.5E-02	0.823	2.0E-02
GDP in 1960 (log)	0.945	-1.2E-02	0.685	-9.0E-03
Initial spending per secondary student in USD logged	0.836	7.1E-03		
Real exchange rate distortions	0.736	-1.2E-04	0.082	-7.9E-05
Life expectancy in 1960	0.655	7.5E-04	0.209	8.0E-04
Malaria prevalence in 1960s	0.495	-1.5E-02	0.252	-2.0E-02
African dummy	0.372	-1.4E-02	0.154	-2.0E-02
Fraction Confucian	0.229	4.3E-02	0.206	5.0E-02
Fraction GDP in mining	0.167	4.2E-02	0.124	4.0E-02
Primary schooling in 1960	0.133	1.8E-02	0.796	3.0E-02
Fraction of tropical area	0.116	-9.5E-03	0.563	-2.0E-02
N				76
iterations				50.2m

Note: Dependent variable is average real GDP growth between 1960 to 1996, expressed as a fraction, as in 0.024.

If one compares the above results with the original SDM results, what stands out is the very high inclusion probability (PIP) of initial spending per secondary level student. This variable is likely to be reflecting much of what was captured in the one education variable that emerged as important in the original SDM work, namely the primary level GER in 1960. The secondary spending variable appears to have pushed the primary GER variable down the ranking. The latter variable came in second after the Asian dummy in the original SDM results[7], whilst in the above table the PIP for the primary GER is just 0.133. Early spending per student at the secondary level appears to be a far better proxy for a well organised education system and general commitment to education, than the primary GER. The results point to an increase in the initial spending per secondary student of 50% being associated with a 0.3 percentage point improvement in annual GDP growth. It should be noted that the coefficients for the 1960 GDP per capita and secondary spending variables (variables 2 and 3 in the list) have opposite signs. This is remarkable partly because the two variables display a high and positive correlation with each other, of 0.718. We can interpret the results as meaning that countries with especially low incomes but which also displayed exceptionally high per student spending in the initial period, were those countries most likely to grow fast.

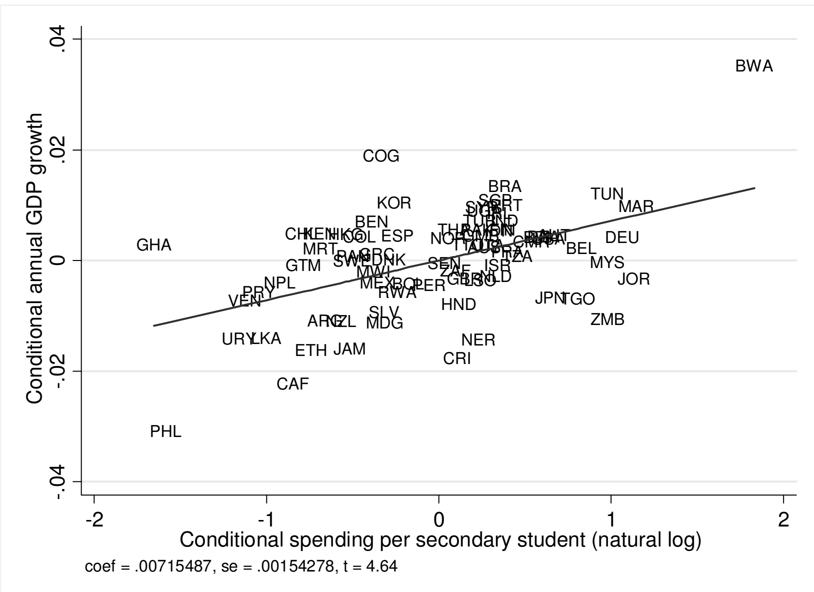
An added variable plot graph can assist in understanding the role played in a regression by specific observations, or countries in our case, when a new variable, such as spending per secondary student, is introduced. In Figure 57, the vertical axis represents the residual, or estimated error term, when the growth variable used for the Table 35 results is regressed on seven key explanatory variables listed in Table 35, but

[6] The threshold was a PIP, or posterior inclusion probability, of 0.097, derived by dividing an assumed 7 variables for an optimum model by the actual 72 explanatory variables used in the BACE process, the 72 being comprised of 67 original SDM variables and the five new education ones. The PIP refers to the probability that the variable in question would be included in the optimum growth regression.

[7] In the original SDM model, the PIP for the primary schooling variable was 0.796, against 0.133 in the results shown here.

not the secondary spending variable, using a simple OLS regression. The seven key variables are the first seven listed in Table 35, excluding the secondary spending variable, plus the primary GER variable. The horizontal axis represents the residual when the secondary spending variable is regressed on the seven key explanatory variables. The utility of the graph can be explained through reference to Botswana (BWA). This country lies to the right because its secondary school spending was high in the 1970s when controlling for a number of factors. It also carries a high value against the vertical axis because there is considerable positive growth which is not explained by a typical growth regression, in this case a regression that excludes secondary school spending. What all this suggests is that the greater predictive power brought about by including the secondary spending variable would rely strongly on a country like Botswana, where secondary spending would 'predict away' much of the residual growth, thus making Botswana less of an outlier in the graph. Similarly, Philippines (PHL), with negative residuals on both axes in Figure 57, would end up with a far smaller residual because of the introduction of the secondary spending variable in the regression. However, the introduction of the secondary spending variable, which raises the adjusted R^2 value from 0.682 to 0.759, would reduce the residuals of many countries to some extent. In fact all countries to the top right and the bottom left of the intersections of the two zero origins in the centre of the graph would see their residuals reduced. The influence of, say, Botswana on its own would not be large. If one excluded Botswana from the regression that included the secondary spending variable, the positive coefficient on this variable would be reduced only by around a fifth. If Philippines were also excluded, the impact would be larger. Overall the magnitude of the secondary spending coefficient would be cut by a half. Thus spending per secondary student in the 1970s emerges as a predictor of growth in the subsequent decades, largely because it explains, in a statistical sense, what occurred in a number of developing countries. From a policy angle one can speculate that ensuring that the secondary level attracted sufficiently talented teachers, through adequate salaries, and was sufficiently equipped with books and other materials, was important for the overall economic development process.

Figure 57: Added variable plot for spending on secondary students



10.6 Guidance for a country's policymakers

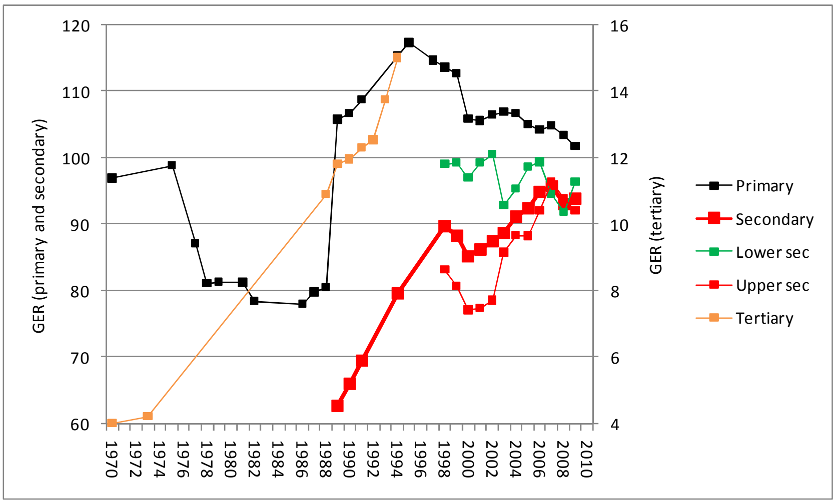
The preceding sections have identified patterns in the cross-country data relating to the prioritisation of the various sub-sectors of the education sector. Examining where a specific country lies in relation to these patterns can strengthen or challenge prevailing understandings amongst policymakers. In this section, South Africa's position is examined, partly with a view to checking the utility of what has been presented above. However, considerable attention also goes towards examining the reliability of the UIS statistics and recalculating these statistics where necessary. In fact, it will be seen that getting the basic facts, or accounting, right, and possibly arguing that firmly entrenched statistics are seriously incorrect, can represent a large portion of the work that must be done by anyone wishing to advice policymakers.

Gustafsson (2012b) points to serious problems in the way the UIS and others have calculated the gross enrolment ratio (GER) values for South Africa, relating to the incomparability of official population estimates and official enrolment estimates. The suggestion made is that accurate GER values should be calculated from one source, household survey data. This calculation is done below. Moreover, key spending values are recalculated, and here too it is found that the UIS values are problematic. All this raises the question of the extent to which UIS values in other countries are reliable, and thus the validity of the kind of cross-country analysis presented in previous sections. In defence of the UIS data it can be argued that the patterns seen in the previous sections seem sufficiently non-random and intuitively right to warrant faith in the data. One can assume that to some degree country-specific errors are randomly distributed and thus cancel each other out, producing aggregates that are usable. Gustafsson (2012: 6), in comparing GER-type errors in South Africa to those found elsewhere, concludes that those for South Africa are particularly large, which also counts in favour of the overall reliability of the UIS statistics.

Trends for all the UIS variables out of the 11 UIS variables that have received attention above and for which values existed for South Africa, are shown in the next three graphs[8].

[8] An anomaly that applies to very few countries is found if one compares one of the trends, that for tertiary GER in Figure 58, to several issues of the *Global monitoring report* produced by UNESCO (for instance UNESCO, 2005). Whilst the UIS data described in section 10.1, and downloaded in 2012 off the UIS website, contain no tertiary GER values for South Africa beyond 1994, the UNESCO reports do have post-1994 values, though the *Global monitoring report* is supposed to be based on the same UIS data source as that used for the previous sections, and in the case of virtually all countries the reports do mirror the online data exactly. Specifically, the *Global monitoring report* issues for the years 2003 to 2010 all have recent tertiary GER values for South Africa (though the 2011 and 2012 reports have no tertiary GER statistics at all, for South Africa or any other country). The tertiary GER values for South Africa published in the UNESCO reports are in line with figures presented below, for instance 15.2 for 2000 and 15 for 2007. South Africa is not the only country where this strange gap in the UIS data occurs with respect to the tertiary GER, yet this problem seems highly unusual. An analysis of the 2007 tertiary GER values published in the 2010 *Global monitoring report* reveals that only Ecuador displays a similar anomaly of a value in the report, but not the UIS database. One could speculate that statistics were not submitted by the South African government to UNESCO through the usual route and that UNESCO, realising the importance of South Africa, then used publicly available reports to construct the GER values. Such patching of the data would be relatively easy. However, if one considers that a few very important countries, with much publicly available data, notably Germany and Singapore, have missing tertiary GER values for recent years in both the *Global monitoring report* and the UIS database, it seems curious that UNESCO should produce a special patch for South Africa

Figure 58: South Africa gross enrolment ratio values



Source: Own calculations using data from UIS online querying facility, downloads from September 2012.

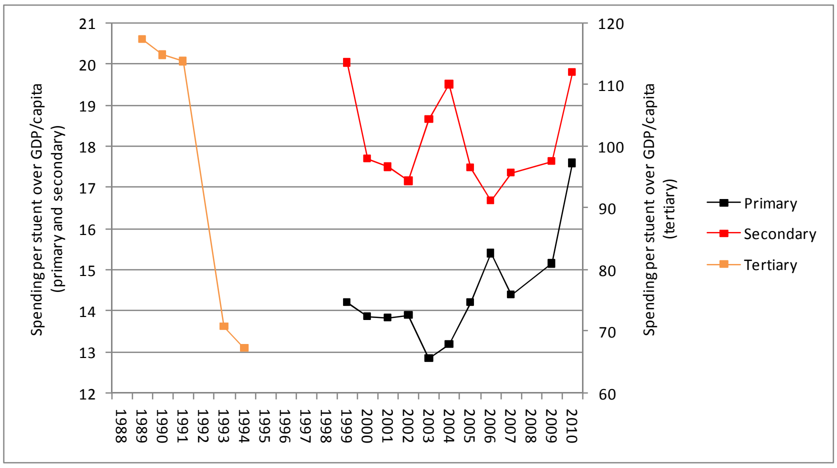
According to the above graph, 1989 is a watershed year in terms of the quality and availability of GER values. Before that year there are hardly any tertiary GER data points, there are no secondary GER values, and the primary GER trend looks suspicious. One might have expected some negative effect on primary enrolments of the civil unrest occurring in the 1976 and 1988 period, but the likelihood that these troubles cut primary enrolments by a fifth, as suggested by the graph, and that a recovery should have occurred so quickly, between 1988 and 1989, seems unlikely. In fact, a historical enrolment dataset compiled by Crouch (1999), using a variety of pre-1994 sources, points to a steady increase in absolute primary level enrolment numbers between 1981 and 1994, with increases each year in the range of 0.4% and 6.6%. The precipitous dip in primary enrolments after 1976 seen in the above graph is moreover not seen in the form of a kink in the later attainment statistics amongst adults in Louw, Van der Berg and Yu (2006: 30) or Case and Deaton (1999: 1055). The large dip in the primary GER seen in the UIS dataset for South Africa is almost certainly nothing more than a manifestation of the serious measurement problems applicable to South Africa during the pre-1994 period, when officially the country was considered to be split into several independent republics and data collection processes were fragmented. The rather sharp decline in the primary GER for the *post*-1994 period seen in the above graph is also deceptive. This matter receives some attention below.

An analysis using fresh data presented below will indicate that the primary and secondary spending ratios seen in Figure 59 are more or less right, though the real trend is almost certainly less erratic than what is seen in the UIS data[9].

outside the normal data collection procedures. The absence of proper metadata reports for the UIS data, of the kind one finds for the Penn income data, obviously makes it difficult to resolve questions of this kind. The answer is probably only obtainable by interviewing UIS staff.

[9] The problem in relation to South Africa's tertiary GER discussed in the previous footnote does not apply in the case of the tertiary spending ratio, though here too there are no post-1994 values. The *Global monitoring report* has never published any tertiary spending statistics, for any country, so here it seems clear that what is in the online database represents all the data there is.

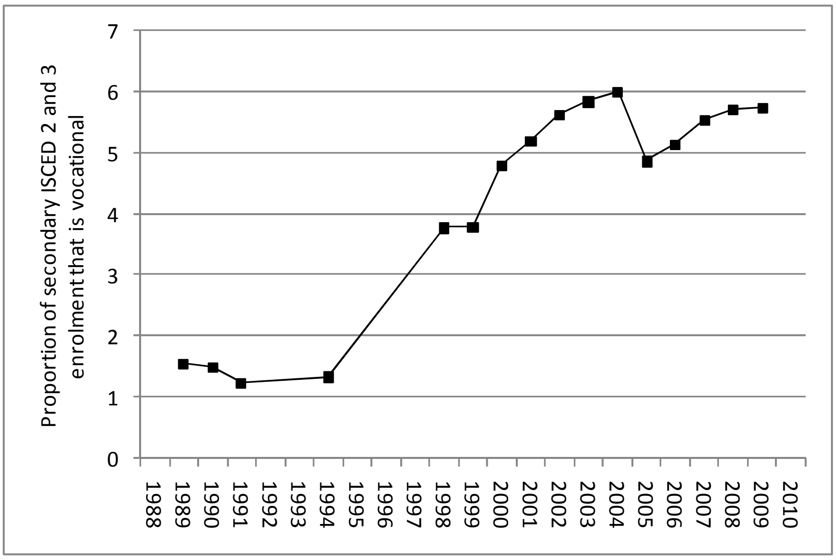
Figure 59: South Africa UIS education expenditure indicator values



Source: Own calculations using data from UIS online querying facility, downloads from September 2012.

The proportion of secondary level enrolments that is vocational has grown over the long term, according to the following graph, though the trend since around 2000 has been uneven. This UIS pattern agrees roughly with what is seen in official publications (for instance South Africa: Department of Higher Education and Training, 2013), though here the absence of UIS metadata poses a serious problem as vocational enrolment statistics can be derived in two very different ways. They can represent actual people enrolled at some point in the year (headcounts) or such people converted to full-time equivalent students. It appears as if the UIS statistic reflected below is derived from headcounts and not full-time equivalent students.

Figure 60: South Africa UIS vocational secondary education enrolment values



Source: Own calculations using data from UIS online querying facility, downloads from September 2012.

The enrolment and spending indicator values for all three levels of the education system in South Africa are now recalculated, going back to around the middle of the first decade of the 21st century. For education planners to understand whether current trends are moving in the right direction, data for just over five years are generally enough. Gustafsson (2012b) points out that dividing official enrolment values by population values results in under-estimates of the GER because population estimates appear to be over-estimated by around 14%. Table 36 presents what can be considered more or less accurate GER figures, derived from just one data source, the sample-based General Household Surveys (GHS) (Statistics South Africa, 2012), meaning the problem of a numerator that is not comparable to a denominator is avoided. The Table 36 GER values are indeed 5 to 10 points higher than the GER values of the UIS seen in Figure 58. The age ranges and school grades used for the Table 36 ratios are the same as those used by the UIS.

Table 36: South African gross enrolment ratios from household data

	2003	2006	2009	2010	2011	2006-2011 slope
Primary (7-13)	112	115	115	115	116	0.1
Lower secondary (14-15)	107	104	109	111	115	2.2
Upper secondary (16-18)	93	99	95	93	97	-0.7
Secondary (14-18)	99	101	101	100	105	0.4
Tertiary (19-23)	15	13	14	13	15	0.2

Note: Values in brackets refer to the age range used for the denominator.
Source: Own calculations using General Household Survey datasets (Statistics South Africa, 2012).

From 2009, the GHS started asking what grade each child was attending. Responses to this question were used in the calculation. However, before 2009 the grade attended had to be derived indirectly through responses to two questions, firstly what the highest grade was that the child had successfully completed and, secondly, whether the child was enrolled in a school. This would result in an inflation of a few points in the pre-2009 GER values at the primary level, due to ambiguities in the data in relation to pre-schooling[10]. Even if one takes these measurement problems into account, it remains clear that the GER values for South Africa are in fact higher than what is reflected in the UIS database, and that with the exception of the lower secondary level, they have been relatively stable in recent years. The lower secondary level (grades 8 to 9) ratio has seen an average annual increase of around 2.2% in the 2003 to 2011 period.

The next table presents statistics needed to calculate fresh spending ratios for South Africa, for the years 2006, 2009 and 2012. The process of obtaining these ratios illustrates the complexity of the exercise and, by implication, how easy it is for spending ratios in the UIS data collection system to be corrupted. Non-bold values in

[10] If a child had successfully completed grade 2 and was currently enrolled in a school, it was assumed the child was attending grade 3. The indirect approach could also be followed in the case of the 2009 data and if this was done, the secondary GER would be 100, against 101 if the direct method were used. At the primary level, however, the two approaches yielded more different ratios: 119 if the indirect approach was used and 115 in the case of the direct approach. This is not surprising if one considers the problems posed by the indirect approach when grade 1 enrolments are calculated. The approach dictates that a child enrolled in a school and with a response to highest grade attained being 'no schooling', would be considered a grade 1 pupil, but this child may also be enrolled at a level below grade 1 in an institution considered a school by the respondent, even if the strict definition would be a pre-school or a crèche. Grade 1 enrolments, and hence the primary GER, are thus over-estimated if the indirect approach is used, though this problem would have been smaller in, say, 2003 than 2009 as enrolments below grade 1 would have been lower in 2003 than in 2009.

Table 37 are source values used to calculate the values in bold. GDP per capita values in the table are actually drawn from two different sources: GDP from official Statistics South Africa (Stats SA) reports is divided by official Stats SA mid-year population estimates (Statistics South Africa, 2010, Statistics South Africa, 2011, and so on)[11]. Total current spending on schools is taken from National Treasury publications, and only figures for the budget programme dealing with public ordinary schools are considered (http://www.treasury.gov.za). Thus the effect of public funding on private school students, which would diminish the average spending per student (by around 2% only), and the effect of public funding on special needs students, which would increase average spending per student slightly, are ignored. The strong possibility that some pre-school spending, or spending on grade R, a grade below grade 1, is erroneously included in the budget programme on public ordinary schools, and not in the separate budget programme created for grade R, is a further complexity that is not explicitly dealt with. Moreover, it is assumed, for instance, that for the 2006 school year, which lasts from January to December, one can use figures for the 2006/07 financial year, which lasts from April 2006 to March 2007.

Spending on the primary level is seldom accurately separated from spending on the secondary level partly because per student spending on these two levels is rather similar and is driven by the same set of policies, and partly because there are many schools with a mix of primary grades (1 to 7) and secondary grades (8 to 12)[12]. For Table 37 a ratio of 1.12 for per student spending at the secondary level, relative to the primary level, was used, based on analysis put forward in an unpublished report of the national education authorities (South Africa: Department of Education, 2006). This report contains what is perhaps the most in-depth analysis available of how costs at the secondary and primary levels differ, using to a large extent differences seen in the payroll data and differences in pupil/teacher ratios. Though the same set of policies governs primary and secondary school funding, secondary schools tend to be funded somewhat more, in per student terms, because of the way the teacher allocation norms treat the secondary school curriculum and because secondary schools tend to be larger, which helps them secure more publicly paid administrative and support staff, and higher salaries for school principals.

[11] The possibility that the inaccuracies in the latter figures, discussed above, cause inaccurate GDP per capita figures is not be discussed here, as this would be a complex matter beyond the scope of this dissertation

[12] National Treasury publications do in fact report different spending figures for the primary and secondary levels, within two sub-programmes falling under the public ordinary schools programme, but this is largely done to satisfy International Monetary Fund reporting demands, not because this breakdown is used by local analysts. In fact, some basic analysis, by province, of the published primary and secondary spending values against official enrolments will reveal that there must be serious inaccuracies in this split of the public ordinary schools totals.

Table 37: Calculating South African per student spending ratios

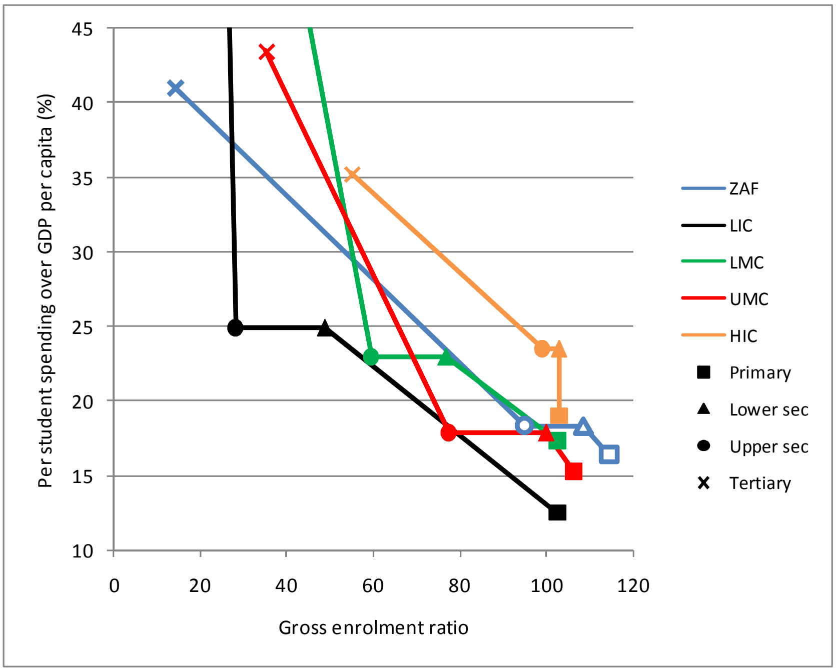
	2006	2009	2012	2006-2012 slope
GDP per capita at current prices (Rand)	37,295	48,791	63,115	
Total current spending on schools (million R)	59,934	93,527	121,438	
Enrolment primary (grades 1 to 7) (thous.)	7,078	6,917	6,744	
Enrolment secondary (grades 8 to 12) (thous.)	4,400	4,263	4,406	
Ratio of secondary to primary unit cost	1.12	1.12	1.12	
Per student spending primary	**5,000**	**8,012**	**10,415**	
Ratio to GDP per capita	***0.13***	***0.16***	***0.17***	*0.005*
Per student spending secondary	**5,579**	**8,940**	**11,621**	
Ratio to GDP per capita	***0.15***	***0.18***	***0.18***	*0.006*
Subsidies to universities (Rm)	10,986	15,297	20,903	
Subsidies to university students (Rm)	926	1,444	3,875	
Enrolment universities	741,000	837,779	909,716	
Per student spending university	**16,076**	**19,983**	**27,237**	
Ratio to GDP per capita	***0.43***	***0.41***	***0.43***	*0.000*

Turning to public spending per student at the tertiary level, here much of the complexity relates to public funding for student loans and bursaries occurring through the National Student Financial Aid Scheme (NSFAS), captured in the line ‘Subsidies to university students’ in Table 37. As for pre-tertiary education, a key source for tertiary spending are the publications of National Treasury. This source is used for the fairly straightforward line ‘Subsidies to universities’. A problem with the National Treasury figures for subsidies to students is that these figures, expressed as transfers to NSFAS, do not distinguish between funds that flow to tertiary or university students, and funds that flow to non-tertiary vocational college students. By 2012, the proportion of NSFAS spending flowing to college students seemed to have reached 25% (South Africa: Financial and Fiscal Commission, 2012: 45; South Africa: National Student Financial Aid Scheme, 2012: 53). In 2006 the figure was 0%, and in 2009 around 17%. Though NSFAS is able to provide loans and bursaries in excess of its receipts from National Treasury, as loan repayments get recycled into new subsidies, only the portion of National Treasury’s contribution considered to be flowing to non-college students was counted. The would be in line with the UIS approach of counting only public funding, meaning revenue from private loan repayments should be ignored. The approach followed gives one what is probably the closest one can get to counting just subsidies, and not loans, in the sense that in a context of stable enrolments and prices, the National Treasury contribution would more or less amount to subsidies excluding loan amounts, as loans would be funded by the debt repayments of former students. The sum of the subsidies to universities and subsidies to students lines of Table 37 agrees broadly with other estimates of the total public funding of university students (South Africa: Financial and Fiscal Commission, 2012: 45; De Villiers, 2009). What seems to be the most recent UIS guide on how to calculate several (but not all) the UIS indicators, UNESCO (2003), indicates that the denominator for the tertiary spending ratio should be headcounts and not full-time equivalent students. This is important as if one uses full-time equivalent students, the denominator drops by around 35%, meaning the ratio would rise by around 50% (South Africa: Department of Higher Education and Training, 2013)[13].

[13] One calculation that was not done was to remove the public spending associated with capital investments from overall spending values, to produce recurrent spending figures, the ideal for the

The previous two tables present enrolment and spending indicator values for South Africa which we can consider sufficiently reliable for a meaningful comparison to global trends, using the Figure 37 graph. For South Africa, 2009 values are used and global values are on average from 2007, as in Figure . The graph is magnified below to provide a focus on values of relevance for South Africa (ZAF).

Figure 61: Educational level prioritisation in South Africa 2009



Source: Own calculations using data from UIS online querying facility, downloads from September 2012.
Note: The expenditure levels used for lower and upper secondary are those that apply to secondary schooling as a whole.

The spending levels in South Africa are more or less like those in other upper middle income countries, according to the graph. The graph indicates that South African primary and secondary spending levels are marginally above the average, whilst the tertiary spending level is marginally below. The more interesting comparison relates to enrolments. Here South Africa lies considerably above the average UMC level when it comes to schooling, and considerably below this level when it comes to tertiary enrolments. In fact, the primary and lower secondary GER values for South Africa lie to the right of the corresponding averages of all development categories. At the tertiary level, South Africa's GER of 14 is closer to the low income average of 9 than the lower middle income average of 25, let alone the upper middle income average of 35.

Figure 62 below offers a more continuous view of the global patterns and South Africa's position within them. This is a reproduction of Figure 43, meaning development trajectories are depicted not primarily on the basis of income, but on the basis of a factor analysis of the education indicators. South Africa's actual position is indicated by the solid triangles, whilst the positions one might expect South Africa to

indicator according to UNESCO (2003). No analyst appears to have done this and doing this would involve examining the financial records of all the individual universities.

be at are indicated with the non-solid triangles. South Africa was not one of the 112 countries for which sufficient UIS data existed for inclusion in the factor analysis, thus South Africa's expected positions along the curves were calculated by using Figure 42 to translate South Africa's income value to the z-score discussed earlier.

Figure 62: Average pathways with South African points

Source: Own calculations using data from UIS online querying facility, downloads from September 2012.
Note: The points labelled 'LIC' etc. are positioned where they represent the median z-score per development category produced by the factor analysis.

How might the above graph assist the South African policy debates? It can reinforce the notion that South Africa is at the lower end of the upper middle income category, meaning that a simple comparison to UMC averages can be deceptive. In fact, along several of the curves South Africa is closer to the LMC median point than the UMC median point. But above all the graph underlines a fundamental structural problem with the South African education system, namely that relatively speaking it has over-prioritised schools and under-prioritised tertiary education, both with respect to enrolment numbers and per student spending. To some extent, the Sub-Saharan Africa problem of an under-sized tertiary sector exists in South Africa too, though South Africa does not experience the problem of an extremely high spending ratio. At the tertiary level, spending per student in South Africa seems a bit low. But the key problem is that tertiary enrolments should be around 70% higher than they were in 2009.

In Table 38 below, the global 'norm' is represented by the non-solid triangles in Figure 62 above, or where South Africa perhaps should have been in 2009. The question is whether overall this would have implied a more or less costly education system, relative to the actual situation in 2009. We can use the enrolment and spending indicator values to calculate the overall public cost of the education system relative to GDP, if we introduce just one statistic, namely each age-specific population group's share of the total population. This is provided in the last column. For instance, youths aged 19 to 24, the ages used for the tertiary GER by the UIS, accounted for 10.1% of the total population in 2009. Clearly, South Africa has not achieved a less costly education system by over-prioritising primary and secondary education relatively to tertiary education. Total spending came to 5.3% of GDP in 2009 (if one considers just the main education sub-sectors listed in the table), and it would have come to around 4.4% of GDP had South Africa's enrolment and spending patterns been closer to global trends. One could use the figures in the table to point out that if a more 'normal' degree of prioritisation towards tertiary education were to be achieved in South Africa, this would raise public spending in this level from 0.6% of GDP to 1.1%, but to 1.0% if 2009 per student spending ratios were maintained. To free up the additional 0.5% (or 0.4%) of GDP needed for this, within the existing budgetary envelope, it might be possible to reduce the schooling GERs over time through quality-related efficiency gains, specifically reductions in the primary level GER through less grade repetition. In fact, moving from the actual primary GER of 115 to a more efficient, yet attainable, GER of 104, would free up around 0.3% of GDP, not much less than the percentage of GDP needed to expand the tertiary level. In fact, this type of solution may be missed by policymakers because primary GER values for South Africa published by the UIS, but also others, tend to be under-estimates.

Table 38: Total spending patterns for South Africa

	Actual			Global 'norm'			
	GER	Exp. per cap./ GDP per capita	Total exp./ GDP	GER	Exp. per cap./ GDP per capita	Total exp./ GDP	Fraction of total pop.
Primary	115	16	2.8	104	13	2.0	0.148
Lower secondary	109	18	0.8	95	14	0.6	0.043
Upper secondary	95	18	1.1	69	14	0.6	0.064
Tertiary	14	41	0.6	24	46	1.1	0.101
Overall	82	18	5.3	74	17	4.4	0.356

Source: Stats SA mid-year population estimates broken down to single ages as explained in Gustafsson (2012b) were used to calculate the proportions in the final column.

New analysis can provide a needed reinforcement of existing analysis and perceptions, and it can assist in overturning misperceptions and 'myths'. What contribution may the above analysis make in the case of South Africa? Amongst analysts, the point has been made before that there is an under-prioritisation of post-school education. This is done by, for instance, Gustafsson (2011), using cross-country comparisons but in a manner that differs to what was presented above. Crouch and Vinjevold (2006) have argued that Southern Africa is an unusual region in its combination of high enrolment ratios at the primary and secondary levels and exceptionally poor performance in international tests. They speculate that the region has perhaps over-prioritised enrolment at the cost of quality, a hypothesis that would complement the tertiary under-prioritisation argument, partly because a robust tertiary education sector relies on there being sufficient quality in the schooling sector. The policy documents do reflect a concern with the low levels of tertiary enrolment. For example, as discussed in Part III of the dissertation, the government aims to increase post-school enrolments, including university enrolments, by a factor of 4.7 by 2030. However, this target is unrealistically high and can be considered a reflection of the fact that the necessary planning has not occurred. With respect to the high secondary level enrolment values in South Africa, there is no plan to reduce these, as one might expect. However, there appear to be risks inherent in the policy uncertainty governing this area, and intermittent signals that enrolments should grow even further at this level. A 2009 set of targets released by the Presidency referred to 'enrolment ratios' of 95% by 2014. If this was simply taken to mean the gross enrolment ratio, then this 95% had already been reached by 2009 at the upper secondary level (see Table 36). However, in the South African context the target is more likely to have been interpreted as the percentage of youths reaching Grade 12 at school. In 2009, this was around 60%, which could be broken down into around 40% successfully completing Grade 12 and a further 20% participating in the grade but not passing (South Africa: Presidency, 2009: 23; Gustafsson, 2011). A clearer but different target emerges from the 2011 sector plan for basic education, where it is envisaged that by 2019 60% of youths would be successfully completing Grade 12 at a school, whilst the remaining 40% would be achieving something equivalent in the vocational training sub-sector (South Africa: Department of Basic Education, 2011b: 175). The 2012 national development plan, on the other hand, puts forward as a target twelve years of compulsory schooling by 2030 (South Africa: National Planning Commission, 2012: 296). The very strong focus on completing Grade 12, or at least twelve years of education, a commitment which could result in even further growth in the secondary schooling level, needs to be understood against the fact that there is no national qualification below the Grade 12 level, so amongst policymakers and the public there is a sense that someone who has not completed twelve years of schooling is essentially uneducated. This represents an unrealistic and extreme position that ignores the fact that even in some high income countries there are youths who enter the labour market with less than twelve years of schooling. For instance, in recent years around 10% of youths in the United States have completed just grade 11 or a lower grade[14].

Turning to per student expenditure, the perception that spending is high at the school level relative to what is found in other countries is commonly found in both the analytical and policy texts. For instance, the 2011 sector plan quotes figures indicating that USD PPP spending per secondary student is 2.9 times the level found in Latin America (South Africa: Department of Basic Education, 2011b: 28). If one uses per student spending over GDP per capita, as is done in the above analysis, the figure of 2.9 is reduced to 1.4. Clearly the magnitude of the difference is influenced by the measure one uses. The policy challenge, if one takes into consideration analyses of what drives costs in schools, seems to be to ensure that teacher pay is not allowed to rise uncontrollably and to crowd out other priorities. Moreover, the structure of the teacher pay system needs to become more oriented towards incentivising improved educational outcomes (Gustafsson and Patel, 2009; South Africa: Department of Basic Education, 2011b: 16). One way of explaining to policymakers and teacher unions that there is something wrong with the recent school spending trajectory is to refer to the fact that per student spending over GDP per capita ratios have been rising, for instance by 0.5 points a year for the primary ratio (see last column of Table 37) when the general trend for middle income countries is for these ratios to fall (see Figure 49).

[14] Specifically, a table available on the United States Census Bureau site at http://www.census.gov/hhes/socdemo/education/ indicates that for several age groups between 25 and 39 around 10% of the population had completed grade 11 or a lower grade and nothing else in 2012 (viewed October 2013).

11 CONCLUSIONS FOR PART II

Part II of the dissertation is the part most focussed on extracting new findings from the data. In this task, a relatively well-known but arguably under-utilised data source, the country-level education data of the UNESCO Institute for Statistics (UIS), was used. The UIS enrolment data have been used fairly extensively, by for instance Lutz, Cuaresma and Sanderson (2008), to fill in gaps where country-specific household data were not able to provide educational attainment values, but the kind of analysis of cross-sectional and historical enrolment and expenditure trends provided in this part of the dissertation is new. As speculated in section 10.1, a reason why the UIS data have not been subjected to more comprehensive analysis of this nature could be that obtaining a usable dataset off the UIS online data querying facility is cumbersome.

The central question for this part of the dissertation was: *What new patterns can be found in the historical country-level enrolment and per student spending data to inform policymakers on how to prioritise the various levels of the education system?* It should be emphasised that comprehensive and reliable guidance to policymakers on how to identify and resolve imbalances in the spread of resources and policy emphasis across the levels of the education system has been scarce in the economics and education planning literature, despite the obvious importance of the topic. As will be seen, the patterns observed partly validate understandings policymakers and analysts already held, partly challenge some of these understandings, and to some extent offer a few complete surprises.

That there are in fact sufficiently robust patterns in the UIS data that relate to country development is perhaps most strikingly seen in the factor analysis described in section 10.3. Of the 194 countries with at least one value since 2000 with respect to seven key indicators, four gross enrolment ratios for four levels of the system, and three expenditure values for three levels of the system, 112 countries had all seven values. For these 112 countries, factor analysis allowed for the generation of a single z-score summarising relatively well the trajectory of country development. Four of the variables displayed uniqueness scores indicating particularly strong correlations. This appeared to support the viability of depicting comparable convex or ‘boomerang’ curves, each representing the cross-sectional configuration of enrolment and spending across education levels of a single country or a group of countries (see for instance Figure 61), as well as ‘development routes’ for each level of the education system (see for instance Figure 62). It has been argued that these graphs represent devices that policymakers may find useful for describing and comparing education systems and specifically prioritisation across levels.

The historical analysis presented in section 10.4 has confirmed that there has been a striking increase in enrolments at all education levels in the decades since 1970. This increase clearly exceeds what one may expect in a normal country development trajectory where particular levels of development correspond to particular enrolment patterns. It is this trend which in many ways inspired Pritchett’s (2001: 368) hypothesis of massively inefficient over-enrolment across the world. The analysis presented above found that the enrolment expansion trend clearly holds even when changes in the classification of country development status over time are taken into account. Importantly, the trend has been accompanied by a large fall in per student expenditure relative to GDP per capita, at least at the secondary and tertiary levels, even within specific country development categories. For instance, for middle income countries the average gross enrolment ratio (GER) at the tertiary level more than tripled, whilst per student expenditure relative to GDP per capita was by the 2000s a fifth of what it had been in the 1970s. An attempt was made to disentangle the influence of rising income and the influence of historical forces on enrolment increases and it was found that the surge in secondary level enrolments occurred largely in the 1980s and 1990s, whilst that for tertiary enrolments occurred largely in the 1990s. What the data cannot tell us is what these historical forces were, but one can speculate that changes in technology pushed up the demand for, say, tertiary education, whilst the political emphasis on universalising primary school education would have increased the number of young people demanding secondary and then tertiary education.

The historical trends beg the question of whether the same increases in GDP per capita could have been achieved with fewer structural changes in education, in other words with less expansion of the education system and without falls in per student spending relative to GDP. This question is essentially about the relative strengths of screening effects and productivity enhancement effects. We can be rather certain that the latter effects have occurred, on the basis of the work of for instance Vandenbussche, Aghion and Meghir (2006), discussed in section 9.2 above. But could a large portion of the expansion, perhaps most of it, have been a question of 'education for show', or the effect that Pritchett (2001) worried about? It is instructive to keep in mind a few empirically informed magnitudes. The interpretation in section 9.2 of the models of Vandenbussche *et al* (2006) pointed to an improvement in the growth rate, in terms of per adult output, of 0.1 percentage points arising from a 2 percentage point improvement in tertiary attainment levels amongst adults, from 13% to 15%. Table 30 pointed to expansions in the tertiary GER that would roughly correspond to growth improvements of 1.1 percentage points for middle income countries and 1.9 for high income countries between the 1970s and 2000s. Average annual improvements in GDP per capita between the 1970s and 2000s were slightly negative for high and upper middle income countries, and effectively zero for lower middle income countries, even if one excludes the especially poor growth figures of 2009. These are trends seen in the Penn income data used in the foregoing analysis and obviously ignore the fact that there were some countries that did grow rapidly. Yet at the aggregate level they suggest that much of the enrolment expansion was 'for show', though empirical certainty on this matter may never be achieved given data constraints and the complexity of the dynamics at work.

The analysis in Part II focussed partly on whether there were differences in the education trends between faster and slower growing countries within each country development categories. In general, it was found that faster growers prioritised education more in the sense that they had higher gross enrolment ratios and higher levels of per student spending relative to GDP per capita. One exception to this pattern was that amongst less developed countries, more rapid growth was associated with *lower* values in the per student spending indicator at the tertiary level. In terms of trends over time, the Mingat and Tan (1998: 9) finding, using just cross-sectional data, that there is a large drop in spending on every tertiary student relative to GDP per capita as countries develop was confirmed using longitudinal data, but it was furthermore found that the drop with respect to this spending indicator occurs faster in countries that grow more rapidly. It could be that better governed countries are better at managing unit costs at the tertiary level in a manner that protects the quality of education whilst permitting the enrolment of more students. These countries are thus better at employing a rationalist planning approach as opposed to succumbing to the variety of less rationalist pressures described in section 8 above.

One intriguing finding is the explanatory power that spending on each secondary student in the 1970s, expressed in logged USD PPP terms, has with respect to growth in subsequent years. The semi-Bayesian growth regression of Sala-i-Martin, Doppelhofer and Miller (2004), discussed in depth in Part I of the dissertation, was used to gauge the explanatory power of various education indicators derived from the UIS data. The only indicator that emerged as having a high probability of being included in a growth regression, using the criteria of Sala-i-Martin *et al* (2004), was the abovementioned secondary school spending indicator. In fact, the inclusion probability of this indicator was considerably higher than the inclusion probability of primary school enrolments in the 1960s, the key education indicator considered by Sala-i-Martin *et al* (2004). The analysis presented above points to a 50% increase in initial spending per student at the secondary level being associated with a growth improvement of 0.3 percentage points. This finding runs contrary to some other evidence that per student spending has mattered very little for development. For instance, Hanushek and Woessman (2007: 60) make the latter argument strongly, but using OECD data from a smaller set of mainly high income countries. That secondary schooling indicators can display surprisingly strong conditional correlations with growth has also emerged from the analysis by Lutz, Cuaresma and Sanderson (2008), discussed in section 9.2, though the finding here related to the importance of expanding secondary school enrolments. Indicators on secondary enrolments did not emerge as significant predictors in the semi-Bayesian approach used in section 10.5.

A popular perception that Sub-Saharan Africa has over-prioritised primary schooling, relative to other levels of education, is only partially supported by the analysis presented above. What the UIS data show is that although Sub-Saharan Africa did see large increases in its primary level GERs, up to around 2009 countries in the region were still behind other comparable countries with respect to this indicator. At the same time, over the 1970 to 2011 period, the tertiary GER for the region increased by a larger factor than elsewhere, albeit off a low base. In a similar vein, secondary level enrolments grew faster in the region than elsewhere. The ratio of the primary GER to the tertiary GER dropped in Sub-Saharan Africa to a greater degree than it did in other regions. On the other hand, to support future prioritisation of the secondary and tertiary levels in the region, one could point to the fact that whilst by around 2010 Sub-Saharan Africa had caught up to the rest of the world with regard to primary enrolments, its secondary GER values were around half of those found in comparator countries, whilst its tertiary GERs were on average only one-sixth of what one finds in other countries with comparable levels of income.

One matter that was investigated was whether the correlation between country population and spending on each tertiary student suggested high spending was an economy of scale problem. One might expect a country such as Swaziland with a population of around a million and one small university to experience exceptional pressures associated with fixed costs related to, for instance, libraries and small classes in more specialised subjects. Yet the analysis suggested that rather than population, being in Sub-Saharan Africa was the critical factor driving tertiary education unit costs. This finding suggests that a binding constraint on expansions to tertiary enrolments in the region are solvable inefficiencies within the tertiary sector.

Partly in order to test the utility of the graphs and frameworks developed in the analysis, South Africa’s statistics were compared to global trends. One thing this exercise confirmed was the importance of not taking individual country statistics distributed by the UIS at face value, without considering distortions that may have occurred, in particular due to the incomparability of national enrolment and population statistics. Data problems would of course also bias the global statistics, but one hopes that to some extent country-specific errors would cancel each other out, resulting in global patterns that are sufficiently reliable. Moreover, the errors in the South African values appear to be exceptionally large. Recalculated South Africa statistics were used to confirm what a number of policy documents and analyses have said, directly or indirectly, namely that there is an imbalance between relatively low tertiary enrolments and relatively high secondary enrolments in South Africa. The South African policy is perhaps most confusing when it comes to successful completion of Grade 12, with some public targets advocating twelve years of compulsory education and thus an even higher GER at the secondary level, when what is probably needed is better completion rates with the current GER. Put differently, the bottleneck created by insufficient numbers of secondary school graduates ready for university is implicitly said to be best solved through quantitative expansion when qualitative improvements seem more appropriate. The kind of analysis presented above can thus be useful in pointing out to decision-makers that the key problem in South Africa is not one of an insufficient number of secondary school students.

To conclude, whilst the analysis of Part II has not produced an all-encompassing and empirically informed model of how enrolment and spending, by level, evolve in the country development process, it has, one might argue, brought to the fore patterns that would be informative for policymakers and can be used to persuade experts and voters that cross-country data analysis can be useful and does clarify where structural problems lie and what kinds of changes are possible.

Appendix D

The following table provides values from the dataset used for the main analysis above. Countries are classified as low income, lower middle income, upper middle income and high income countries. The values are the means across the various years for which data were available for that country and that variable. The minimum and maximum years, across all variables, are indicated in the last two columns.

Table 39: Levels of education data by country

ISO code	Country	Cat.	GER prim.	GER LSec.	GER USec.	GER sec.	GER tert.	Voc. 2/3	PSNT	Tert. eng.	Exp. prim.	Exp. sec	Exp. Tert.	GDP cap.	Min. year	Max. year
AFG	Afghanistan	LIC	45	34	15	16	1	4	27		11	30	317	728	1970	2010
BDI	Burundi	LIC	60	21	8	7	1	35		5	18	115	941	391	1970	2010
BEN	Benin	LIC	73	38	15	17	3	7		4	12	23	145	1,063	1970	2010
BFA	Burkina Faso	LIC	34	19	7	7	1	11	32	3	27	69	1,391	718	1970	2011
BGD	Bangladesh	LIC	71	65	35	33	5	1	5	3	6	13	40	879	1970	2010
BTN	Bhutan	LIC	61	60	33	30	4	28	35	13	9	46	222	1,933	1970	2011
CAF	Central African Republic	LIC	72	17	8	11	1	8		4	9	22	446	773	1970	2011
CMR	Cameroon	LIC	95	38	22	22	4	23	7	3	7	48	430	1,825	1970	2010
COD	Democratic Rep. of the Congo	LIC	85	43	24	23	2	32			4	31	362	585	1970	2010
COG	Congo	LIC	118	52	23	50	4	10		2	11	34	618	2,033	1970	2010
COM	Comoros	LIC	91	46	27	28	3	1	20		20	27		1,214	1970	2010
ERI	Eritrea	LIC	49	42	19	24	1	1	18	15	11	25	599	737	1989	2010
ETH	Ethiopia	LIC	45	31	10	17	2	3	10	10	19	42	640	453	1970	2010
GIN	Guinea	LIC	50	32	17	20	3	5		10	12	33	399	833	1970	2010
GMB	Gambia	LIC	60	53	46	19	2	7			16	30	341	901	1970	2010
GNB	Guinea-Bissau	LIC	70	23	16	11	2	13						496	1970	2010
GNQ	Equatorial Guinea	LIC	121	41	10	24	3	10			1	3		4,412	1970	2010
HTI	Haiti	LIC	75			14	1	3						1,719	1970	2009
IND	India	LIC	91	69	40	41	8	1	5		11	21	74	1,525	1970	2010
KEN	Kenya	LIC	101	79	31	35	2	2	18	19	17	25	721	1,118	1970	2009
KGZ	Kyrgyzstan	LIC	104	91	73	91	33	6	11	11		32	24	1,764	1981	2010
KHM	Cambodia	LIC	134	42	17	28	2	3	18	3	6	8	60	939	1970	2010
LBR	Liberia	LIC	63	37	28	19	4	9	28	5	9	67	644	1,025	1970	2009
LSO	Lesotho	LIC	103	45	22	25	2	4	16	1	22	91	1,346	928	1970	2010
MDA	Moldova	LIC	95	91	75	86	34	8	1		37	39	42	1,935	1981	2010
MDG	Madagascar	LIC	113	33	13	21	3	5		7	8	30	346	833	1970	2010
MLI	Mali	LIC	41	34	17	13	2	12		2	27	69	630	686	1970	2011
MMR	Myanmar	LIC	105	52	35	32	5	1		5	3	7	36		1971	2010
MNG	Mongolia	LIC	100	85	69	76	27	7	4	17	14	19	22	2,180	1970	2010
MOZ	Mozambique	LIC	81	21	6	9	1	14		10	15	85	1,409	465	1970	2011
MWI	Malawi	LIC	93	40	15	21	1	1	79	33	11	33	2,029	672	1970	2010
NER	Niger	LIC	31	12	4	6	1	4	3	2	32	89	432	578	1970	2011
NIC	Nicaragua	LIC	92	71	50	42	9	13			11	17	93	2,722	1970	2010
NPL	Nepal	LIC	85	59	25	28	4	3		4	13	12	81	883	1970	2010
RWA	Rwanda	LIC	90	24	13	15	3	28			14	39	1,142	768	1970	2011
SDN	Sudan	LIC	49	43	23	21	2	4						1,490	1970	2009

ISO code	Country	Cat.	GER prim.	GER LSec.	GER USec.	GER sec.	GER tert.	Voc. 2/3	PSNT	Tert. eng.	Exp. prim.	Exp. sec	Exp. Tert.	GDP cap.	Min. year	Max. year
SLE	Sierra Leone	LIC	56	44	26	15	1	5	80	1				883	1970	2011
SOM	Somalia	LIC	20	10	7	7	2	18			46	58	132	671	1970	2009
TCD	Chad	LIC	56	20	11	10	1	5	14		10	26	294	839	1970	2010
TGO	Togo	LIC	105	56	22	26	2	7			10	38	756	1,012	1970	2010
TJK	Tajikistan	LIC	97	92	54	85	19	3	20	10		25	18	1,284	1981	2010
TLS	Timor-Leste	LIC	111	56	38	50	13	5					84	1,132	2000	2010
TZA	Tanzania	LIC	77			4	1	10		13	12	94	1,092	717	1970	2010
UGA	Uganda	LIC	82	26	11	12	2	11	18	8	9	123	942	728	1970	2010
ZWE	Zimbabwe	LIC	95			27	3	2		12			75	328	1970	2010
AGO	Angola	LMC	99	24	14	15	2	16		9		42	76	2,539	1970	2010
ARM	Armenia	LMC	99	94	79	90	28	2	27	7	15	20	11	3,321	1986	2010
BGR	Bulgaria	LMC	102	87	95	93	31	28	2	21	19	19	27	6,182	1970	2010
BLR	Belarus	LMC	99	104	65	94	51	10	21	26	12	31	23	7,654	1971	2010
BOL	Bolivia	LMC	105	96	74	65	26	7		17	14	12	44	3,138	1970	2009
CIV	Côte d'Ivoire	LMC	70	30	15	17	3	10		9	21	144	624	1,511	1970	2011
CPV	Cape Verde	LMC	111	100	67	46	8	13	18	10	17	19	131	1,908	1970	2010
CUB	Cuba	LMC	106	96	80	75	30	29	5	2	34	44	70	8,623	1970	2011
DJI	Djibouti	LMC	35	27	15	14	2	20	16	6	24	33		3,204	1970	2011
EGY	Egypt	LMC	83	96	65	58	19	23	8			50	129	2,824	1970	2010
FJI	Fiji	LMC	108	99	64	74	5	4	7		17	19	135	3,630	1970	2009
FSM	Micronesia (Federated States of)	LMC	185	100	74	136	19	2						3,086	1970	2009
GEO	Georgia	LMC	95	90	70	85	38	4	15	16	15	27	11	3,246	1981	2010
GHA	Ghana	LMC	79	68	26	42	3	4	14	8	10	26	708	897	1970	2011
GTM	Guatemala	LMC	84	51	40	28	7	29		17	8	8	49	5,158	1970	2010
GUY	Guyana	LMC	101	109	64	82	6	7	14	7	13	15	143	2,471	1970	2010
HND	Honduras	LMC	103	66	64	32	10	31		18	12	20	75	3,099	1970	2010
IDN	Indonesia	LMC	106	79	50	43	9	16			12	14	40	2,334	1970	2010
IRQ	Iraq	LMC	95	55	30	42	10	8	7	14	9	13	107	5,814	1970	2009
JAM	Jamaica	LMC	100	96	83	71	12	4	54		12	18	91	7,721	1970	2010
KAZ	Kazakhstan	LMC	109	99	84	95	39	7	34			22	11	7,067	1981	2011
KIR	Kiribati	LMC	123	88	75	41		28			20	88		2,786	1970	2009
LAO	Laos	LMC	103	50	29	27	4	9	35	8	7	19	81	1,208	1970	2010
LKA	Sri Lanka	LMC	101	100	71	65	3	1			9	23	149	2,117	1970	2010
MAR	Morocco	LMC	81	64	31	32	8	5		5	20	66	165	2,338	1970	2011
MHL	Marshall Islands	LMC	109	109	70	82	16	4			23	28	73	7,456	1970	2011
MRT	Mauritania	LMC	58	24	19	17	3	5	18		11	29	148	1,330	1970	2010
NGA	Nigeria	LMC	85	35	29	23	4	15					2,113	1,413	1970	2010

ISO code	Country	Cat.	GER prim.	GER LSec.	GER USec.	GER sec.	GER tert.	Voc. 2/3	PSNT	Tert. eng.	Exp. prim.	Exp. sec	Exp. Tert.	GDP cap.	Min. year	Max. year
PAK	Pakistan	LMC	62	40	24	22	3	3	1		10	23	191	1,730	1970	2010
PHL	Philippines	LMC	109	84	68	69	25		17	14	9	8	15	2,136	1970	2009
PNG	Papua New Guinea	LMC	59	27	3	11	2	19		25	21	116	401	2,046	1970	2009
PRY	Paraguay	LMC	106	76	51	39	13	8	1		10	15	45	3,353	1970	2009
PSE	Palestine	LMC	93	92	69	84	34	1	8	7					1995	2010
ROU	Romania	LMC	101	96	78	84	21	31	10	20	15	14	27	6,662	1970	2010
RUS	Russia	LMC	104	87	90	92	53	13	2				12	10,773	1970	2009
SEN	Senegal	LMC	59	27	12	17	4	6			15	29	184	1,239	1970	2010
SLB	Solomon Islands	LMC	86	52	15	20		21			10	119		1,578	1970	2010
SLV	El Salvador	LMC	98	74	43	43	16	20		12	8	8	14	4,869	1970	2010
SRB	Serbia	LMC	102	99	81	90	49	36	1	16	58	14	41	8,102	1999	2010
STP	Sao Tome and Principe	LMC	134	67	22	37	4	4						1,211	1970	2011
SUR	Suriname	LMC	125	87	55	56	8	35		10				8,499	1970	2009
SWZ	Swaziland	LMC	94	55	34	39	3	2	11	5	11	38	351	2,662	1970	2010
SYR	Syria	LMC	105	80	31	52	14	6			12	20	103	3,117	1970	2010
TKM	Turkmenistan	LMC					17							4,740	1981	2009
TON	Tonga	LMC	113	108	96	97	5	5	33		13	14	100	5,258	1970	2009
TUR	Turkey	LMC	102	92	73	54	17	21		13	9	11	57	6,959	1970	2009
UKR	Ukraine	LMC	104	97	95	95	53	12	7	23			34	5,045	1971	2010
UZB	Uzbekistan	LMC	101	95	105	98	13	15		16				1,659	1981	2011
VNM	Vietnam	LMC	107	85	54	63	8	4		20	19	17	61	1,218	1970	2010
VUT	Vanuatu	LMC	114	47	31	23	4	20	53		14	61	176	4,820	1970	2010
WSM	Samoa	LMC	104	100	74	80	4	3	30	5	10	10	188	5,041	1970	2010
YEM	Yemen	LMC	82	51	37	44	8	1	12					1,951	1989	2010
ZMB	Zambia	LMC	98	51		17	2	6		22	10	96	751	1,430	1970	2010
ALB	Albania	UMC	103	99	61	84	11	19		7	20	6	37	3,347	1970	2010
ARG	Argentina	UMC	109	102	69	69	39	57		9	8	15	20	8,617	1970	2009
ATG	Antigua and Barbuda	UMC	97	113	83	91	16	5			8	12	15	10,533	1970	2010
AZE	Azerbaijan	UMC	101	88	69	84	19	6	49	6	6	16	16	4,187	1981	2010
BIH	Bosnia and Herzegovina	UMC	109	101	81	91	35	35						4,273	1990	2010
BLZ	Belize	UMC	112	78	51	64	13	3	34		15	21	32	6,438	1970	2010
BRA	Brazil	UMC	134	118	88	68	12	10		8	15	14	56	7,385	1970	2010
BRB	Barbados	UMC	106	101	110	93	27	2	26		17	24	100	20,234	1970	2010
BWA	Botswana	UMC	94	88	59	43	4	11	52		14	98	877	5,312	1970	2009
CHL	Chile	UMC	110	97	81	74	26	20		21	12	10	44	6,942	1970	2009
CHN	China	UMC	118	90	51	51	8	9	5		5	22	729	2,013	1970	2010
COL	Colombia	UMC	114	89	66	57	17	20	1	29	12	13	34	4,938	1970	2010
CRI	Costa Rica	UMC	106	92	56	54	21	19		15	12	18	62	8,247	1970	2010

ISO code	Country	Cat.	GER prim.	GER LSec.	GER USec.	GER sec.	GER tert.	Voc. 2/3	PSNT	Tert. eng.	Exp. prim.	Exp. sec	Exp. Tert.	GDP cap.	Min. year	Max. year
CZE	Czech Republic	UMC	106	98	85	91	23	40	12	18	18	22	38	17,843	1971	2009
DMA	Dominica	UMC	111	123	80	86	4	8			15	14		3,713	1970	2010
DOM	Dominican Republic	UMC	109	77	64	50	18	6			6	5		5,582	1970	2010
DZA	Algeria	UMC	98	110	54	47	12	8		9	13	36	331	4,900	1970	2010
ECU	Ecuador	UMC	114	70	53	53	25	19		13	4	8		5,036	1970	2009
EST	Estonia	UMC	100	102	94	97	49	20	13	13	20	31	35	11,640	1981	2009
GAB	Gabon	UMC	154	64	31	36	3	21						11,507	1970	2011
GRD	Grenada	UMC	112	116	94	82	13	9	23	3	9	10		7,630	1970	2010
HUN	Hungary	UMC	100	100	94	88	26	25	17	14	30	15	71	11,926	1970	2009
IRN	Iran	UMC	101	96	69	64	13	7	31	29	12	17	104	8,130	1970	2010
JOR	Jordan	UMC	104	93	74	76	21	6		13	14	27	69	3,946	1970	2010
KNA	Saint Kitts and Nevis	UMC	104	100	85	87	6	4			9	20	77	6,976	1970	2010
LBN	Lebanon	UMC	106	91	70	67	34	13		12	14	4	15	13,099	1970	2010
LBY	Libya	UMC	121	122	97	67	17	17		21			23	17,291	1971	2009
LCA	Saint Lucia	UMC	117	87	74	55	7	5	53		17	31	91	8,484	1970	2010
LTU	Lithuania	UMC	96	98	101	96	49	13	5	19	16	24	31	10,636	1981	2010
LVA	Latvia	UMC	99	96	94	95	42	20	4	11	24	30	27	9,710	1981	2010
MDV	Maldives	UMC	126	104	10	39	13	8	91		16			2,162	1970	2011
MEX	Mexico	UMC	113	105	54	58	16	15		19	11	15	40	9,250	1970	2010
MKD	Macedonia	UMC	97	95	72	82	25	27		17	38	10	61	6,439	1971	2010
MNE	Montenegro	UMC	114	106	82	94	28	31						5,736	1990	2010
MUS	Mauritius	UMC	103	92	74	59	7	4	22		12	19	172	5,179	1970	2010
MYS	Malaysia	UMC	95	93	50	56	17	4	18	22	14	24	130	6,631	1970	2009
NAM	Namibia	UMC	118	80	32	53	6	1	15	3	19	26	116	4,210	1970	2010
OMN	Oman	UMC	65	89	73	41	11	6		9	14	24	116	13,863	1970	2010
PAN	Panama	UMC	108	85	54	62	28	29	6	12	11	15	43	5,979	1970	2010
PER	Peru	UMC	116	96	74	68	24	11	24		7	10	19	5,266	1970	2010
PLW	Palau	UMC	107	101	93	97	39					9	81	15,143	1980	2009
SAU	Saudi Arabia	UMC	93	99	90	84	15	4		7	19	19	165	17,603	1971	2010
SVK	Slovakia	UMC	99	92	83	88	32	35	2	18	16	14	32	13,536	1987	2010
SYC	Seychelles	UMC	108	115	99	90		13			16	30		13,379	1970	2010
THA	Thailand	UMC	93	84	58	43	23	17	1	10	13	19	60	4,357	1970	2011
TTO	Trinidad and Tobago	UMC	101	84	76	73	5	2	45	19	13	19	151	13,271	1970	2010
TUN	Tunisia	UMC	108	106	65	50	13	16	2	10	16	30	128	3,776	1970	2009
URY	Uruguay	UMC	110	109	87	79	31	16		10	8	10	26	6,929	1970	2009
VCT	Saint Vincent and the Grenadines	UMC	115	110	73	63	5	8			19	20		3,346	1970	2010
VEN	Venezuela	UMC	104	83	56	58	27	7		18	7	8	89	8,910	1970	2010

ISO code	Country	Cat.	GER prim.	GER LSec.	GER USec.	GER sec.	GER tert.	Voc. 2/3	PSNT	Tert. eng.	Exp. prim.	Exp. sec	Exp. Tert.	GDP cap.	Min. year	Max. year
ZAF	South Africa	UMC	100	97	86	85	11	4			14	18	97	5,844	1970	2010
ABW	Aruba	HIC	112	111	86	96	29	17	4	23	13	19	32		1999	2010
AND	Andorra	HIC	93	91	70	60	8	7	43		12	11	28		1971	2010
ANT	Netherlands Antilles	HIC	139	122	69	99	23	40	11	33					1972	2002
ARE	UAE	HIC	97	94	76	60	9	2			5	7	21	41,971	1971	2009
AUS	Australia	HIC	105	115	198	139	46	39	15	11	17	16	41	27,412	1970	2010
AUT	Austria	HIC	101	102	96	98	35	39	20	14	20	26	50	27,155	1970	2010
BEL	Belgium	HIC	103	137	126	108	40	45	9	11	17	29	40	25,607	1970	2009
BHR	Bahrain	HIC	110	107	95	86	9	14		8	15	16		22,772	1970	2010
BHS	Bahamas	HIC	102	93	85	91	19	6						21,281	1970	2010
BMU	Bermuda	HIC	99	91	71	84	16	6		12	11	16	28	35,302	1970	2010
BRN	Brunei Darussalam	HIC	118	119	78	78	9	6	1	6	6	18	142	62,215	1970	2010
CAN	Canada	HIC	101	99	106	98	77		22	10			35	27,621	1970	2009
CHE	Switzerland	HIC	93	111	83	96	28	35	11	14	24	30	56	32,597	1970	2010
CYM	Cayman Islands	HIC	102	93	84	89	21		8	8					1972	2008
CYP	Cyprus	HIC	100	99	94	96	29	9		6	17	26	46	12,797	1970	2010
DEU	Germany	HIC	102	101	99	101	41	24			13	27	39	25,318	1970	2010
DNK	Denmark	HIC	99	120	126	113	43	28	1	10	32	30	60	26,613	1970	2009
ESP	Spain	HIC	108	109	127	99	39	20	8	17	15	18	24	19,986	1970	2010
FIN	Finland	HIC	98	101	132	110	54	28		26	22	26	42	23,027	1970	2010
FRA	France	HIC	108	108	113	97	39	23	1	13	15	25	33	24,415	1970	2010
GBR	United Kingdom	HIC	104	101	102	91	36	12	0	9	16	26	57	24,504	1970	2009
GRC	Greece	HIC	100	98	95	88	36	17	9	15	7	12	26	19,085	1970	2009
HKG	Hong Kong	HIC	104	97	69	66	22	5	17	16	10	13	56	21,337	1970	2010
HRV	Croatia	HIC	92	97	83	88	33	37		17			31	10,749	1981	2010
IRL	Ireland	HIC	103	102	121	100	33	11	23	11	13	22	47	20,696	1970	2010
ISL	Iceland	HIC	99	99	118	97	34	22	2	7	23	20	33	27,703	1970	2009
ISR	Israel	HIC	102	97	110	91	40	22	4	18	21	22	45	18,797	1970	2009
ITA	Italy	HIC	102	106	92	82	38	35	2	16	20	24	26	23,342	1970	2010
JPN	Japan	HIC	101	102	101	97	37	14	0	17	20	21	26	25,594	1970	2010
KOR	Korea	HIC	104	100	96	85	47	17		31	15	15	18	12,193	1970	2010
KWT	Kuwait	HIC	99	108	102	84	14	2	30		21	11	103	40,968	1971	2009
LUX	Luxembourg	HIC	97	107	87	77	5	47	26	12	24	25	84	45,205	1970	2009
MAC	Macao	HIC	93	109	73	65	44	12		2	8	11	28	21,424	1970	2010
MCO	Monaco	HIC						19			3	7			1971	2010
MLT	Malta	HIC	104	97	85	81	14	16	8	8	11	20	93	13,363	1970	2010
NLD	Netherlands	HIC	103	131	112	110	41	44	2	9	16	24	62	28,980	1970	2010
NOR	Norway	HIC	100	99	130	103	46	28	3	7	29	22	45	34,047	1970	2010

ISO code	Country	Cat.	GER prim.	GER LSec.	GER USec.	GER sec.	GER tert.	Voc. 2/3	PSNT	Tert. eng.	Exp. prim.	Exp. sec	Exp. Tert.	GDP cap.	Min. year	Max. year
NZL	New Zealand	HIC	103	103	135	98	48	6	15	7	16	17	41	20,945	1970	2010
POL	Poland	HIC	100	99	101	89	32	42	7	14	24	21	28	9,594	1970	2009
PRI	Puerto Rico	HIC	108	89	78	83	79		21		31		66	17,799	1970	2010
PRT	Portugal	HIC	119	114	93	78	29	18	1	21	16	23	35	14,258	1970	2009
QAT	Qatar	HIC	110	100	80	77	17	3		4	8	15	30	73,367	1971	2010
SGP	Singapore	HIC						9	36	29	8	14	56	24,725	1970	2010
SMR	San Marino	HIC	94	97	96	96		14	53	15	23	22			1975	2010
SVN	Slovenia	HIC	100	96	102	92	43	38	1	17	23	24	33	19,190	1981	2009
SWE	Sweden	HIC	101	106	142	104	45	31	3	17	30	25	48	26,345	1970	2010
TWN	Taiwan	HIC												14,823	1970	2009
USA	United States	HIC	98	102	89	92	68		5	7	21	23	24	31,449	1970	2010

Education and country growth models

Part III

How policymakers in developing countries deal with internal and external efficiency issues

---

Martin Anders Gustafsson

CONTENTS FOR PART III

Figures and tables

12 INTRODUCTION FOR PART III

Where Part I of the dissertation has focussed largely on the existing cross-country growth modelling and what this means for policymakers, and Part II has presented some new analysis relating to the prioritisation of the different levels of the education system, Part III focuses on understanding the institutional realities of developing country education systems and the thinking of education policymakers. Part III is important for the dissertation as a whole because if economics is to be employed to improve education policymaking, then a clear understanding of the contexts and mindsets of education policymakers is needed. A key question that shapes Part III is thus the following: *What tends to be the policy thinking and the frames of reference of education policymakers, and what are the implications of this for the conversation between economists and education policymakers?* The economics of education scope in Part III is broader than in the previous two parts of the dissertation. It would have been too limiting to consider just the degree to which policymakers have been receptive to the economic growth literature. Instead, examining how receptive they have been to the full range of economics findings relating to education creates a more complete picture of the people and institutions that economists wanting to influence education must understand.

What are the economics topics which are of particular relevance to education policymakers? There are many ways of answering this question. For instance, Part III of the dissertation sees the economics of education largely in terms of two large concerns: the external efficiency of the education system, or education's contribution to income and development; and the internal efficiency of education institutions, or the ability of these institutions to produce the expected education outcomes. An alternative breakdown is that of Gustafsson and Mabogoane (2012), who see the economics of education broken down into five key analysis areas: labour market rates of return to schooling (in other words, external efficiency); production functions describing how education institutions produce education (internal efficiency); models of teacher incentives; benefit incidence analyses dealing with the equity of public investment in education as well as the equity of rates of return; and cross-country models that compare education systems.

Section 13 below discusses the practical and conceptual hurdles that researchers are likely to encounter if they try to identify the policies of an education system. Partly to create an analytical framework, the policy advice of the economists is also discussed, first with respect to the external efficiency and then the internal efficiency of the education system. Alongside efficiency, educational and socio-economic equity are also dealt with. All this is done in preparation for the analysis of four countries in section 14. Given the relatively strong presence in the literature on education policy (but not really in the policies themselves) of positions that are fairly strongly opposed to those of the economists, a few examples of these positions are discussed. This is done partly to test, at least in a theoretical sense, how well the economics framework holds out against its critics.

Section 14 focuses on the policy positions of the policymakers largely through an examination of selected policies and policy discussions relating to South Africa and three other developing countries: Brazil, Chile and China. The rationale for this selection of countries is explained. What policymakers appear committed to is compared to the policy positions of the economists. Whilst a variety of education policy areas are explored, the area that is of special importance to the dissertation is given prominence, namely the overall gearing of the education system towards national development. The treatment of the four countries could of course have been much deeper, to the extent of becoming a dissertation on its own. However, it will be argued (in the conclusion in section 15) that even an apparently cursive coverage of the education policy dynamics of a few countries of this nature is original and valuable for highlighting policy approaches that should be better known in education policymaking circles. It is also argued that systematic comparisons of country education systems, or even one-country system descriptions within an economic framework, are much rarer than they should be.

Matters of epistemology receive some attention in Part III, specifically how the knowledge of education policymakers is influenced, or not influenced, by economics of education, but also more broadly how empirical evidence and ideology in a sense compete within the policymaking process. What one may mean when one refers to the influence of ideology in policymaking, or in research, is also discussed. These matters are crucial for understanding how economics can make a positive contribution to the education policy discourse. For instance, in using economics to dispel misperceptions where ideology may have clouded views of what is possible, it is important to know where the most serious misperceptions lie. Moreover, a careful examination of what knowledge policymakers need, and the planning and welfare implications of not having the required knowledge, should guide the economists when they decide where to invest their analytical efforts.

13 HOW TO IDENTIFY AND DESCRIBE POLICY PRIORITIES

Part III of the dissertation focuses on how education policymakers in developing countries deal with internal and external efficiency issues. More precisely, what this means is that the policies pursued in a few developing country education sectors are examined, using an economics-oriented framework of internal and external efficiency, both to arrange and evaluate the policy information. The term ‘policy’ is used in a fairly standard manner here, to refer broadly to the guidance provided by documents, mostly (though as will be seen, even by informal understandings), in order to ensure that a public administration behaves in a certain way and pursues certain priorities. The documents providing the guidance could be anything from a national constitution, to a law, to a five-year plan, to a public funding regulation. A search through the literature reveals that there is in fact no widely accepted definition of ‘policy’, though the term may have a specific meaning within one country’s tradition and legal system. This absence of a strict definition is understandable, given the very different institutional make-up of different governments and legal systems. What may be captured in a national constitution in one country may be captured in a regulation in another. Moreover, it is assumed in the policy analysis that follows that although it is preferable to refer to documents, such as laws or plans, it is also possible to consider actions as a proxy for the policy. If policymakers continuously ‘push’ the education system in a particular direction, one can assume they are doing this because there is policy of some sort, possibly formal and documented, possibly more informal, such as the pronouncements of influential politicians, which compels or allows them to do this.

In preparing the ground for the policy analysis of the following section, this section discusses the multi-layered nature of policies, in particular in developing countries, through reference to three key contributing factors: corruption, foreign donor funding and what will be referred to as a culture of institutional ritualism and mimicry. A proper understanding of these matters is vital, it will be argued, for a sufficiently realistic policy discussion. These phenomena are not unique to developing countries, but they tend to be more pronounced in these countries. A narrow definition of policy, for instance a legalistic one, would not be useful in many developing country contexts given that policy takes so many different forms. This section also discusses what economists advise education policymakers, as the economists’ view of optimal education sector policies is used as the framework for the policy analysis of the following section.

The policy recommendations of the economists are relatively easy to identify. They are commonly expressed in the academic literature, as an examination of regular academic articles published in the *Economics of Education Review* during 2011 suggests. Of the 67 articles published in this journal in 2011, 53 included policy recommendations. However, only six of these had policy recommendations relating to developing countries. The World Bank is perhaps the most quoted source of economically-focussed education policy recommendations applicable to developing countries. Key recent publications in this regard are the World Bank’s (2011a) *Education strategy 2020* and the review of best practices for improving the internal efficiency of the education system produced by Bruns, Filmer and Patrinos (2011). Certain UNESCO publications promote the positions of the economists. Notably, the 2005 *Education for All global monitoring report* of UNESCO was informed by background papers of the education economists Eric Hanushek and Martin Carnoy and included substantial coverage of the production function work of the former economist (UNESCO, 2005). The UNDP's annual *Human development report* has promoted the positions of, for instance, the development economist Amartya Sen, with respect to the centrality of human capital in the country development process. For instance, the 2010 report in many respects re-affirms the human capital model, whilst emphasising the need to update it with a clearer treatment of the role of human capital in dealing with new environmental challenges (UNDP, 2010). At the country level, in South Africa a series of working papers of the Department of Economics at Stellenbosch University deals to a large extent with economics of education analysis and associated policy challenges: 9 of 24 papers in 2011 and 11 of 29 in 2010 deal with education[1]. A similar series of working papers, also with a considerable education focus, is maintained by the University of Sao Paulo in Brazil[2]. A recent collection of articles examining basic education questions in Brazil includes articles by 18 Brazilian economists (Veloso *et al*, 2009). Despite these efforts, what stands out is still a large gap between the volume and quality of economics of education papers in developing countries, relative to developed countries. What the policy positions of the economists are is discussed below in this section.

What is far more difficult than identifying the policy positions of the economists, is identifying the policy positions of policymakers themselves, particularly in developing countries. Part of the reason is simply that education policies are not as systemically published as academic articles. It is often difficult to find the education policies of developing countries on ministry websites, for instance. An important tool that to some degree fills the gap for those wishing to compare policies across countries is the online Planipolis facility of UNESCO's International Institute for Educational Planning (IIEP)[3]. This facility provides an archive of national policy documents and plans. To provide an idea of size, there were around 1,700 documents in early 2012, 23 of which were from South Africa.

More insuperable obstacles to effective identification of the actual policies of countries exist, however. Policies, especially in developing countries with weak or non-existent democratic institutions, tend to exist on two levels. On one level, highly publicised and idealistic policies, designed more to satisfy current aspirations than to serve as practical guidance, exist. Inbar (1996: 92) refers to these policies as the outcome of a ritualistic process. On a second level, less publicised and more realistic policies are found which are the policies that implementers actually use. In short, the policy analyst must take care to separate ritualistic documents from documents that actually guide what people do. Where democracy of some form exists, ritualistic policies may be released whose intention is to outdo the promises of a previous government and which are informed by political aspirations rather than empirical analysis of what is doable.

Psacharopoulos (1996: 343), arguably one of the founders of the current economics of education tradition, has argued that 'In the field of education, perhaps more than in any other sector of the economy, politics are substituted for analysis.' Implicit in this observation is the point that education policymakers tend not to appreciate the multi-

[1] See http://ideas.repec.org/s/sza/wpaper.html.
[2] See http://www.cpq.fearp.usp.br/conteudo/seco2.html.
[3] See http://planipolis.iiep.unesco.org.

layered nature of policymaking, and start to believe that the more idealistic policies are actually implementable, and not simply political window-dressing. One can imagine a rational policymaker faced with the need to maintain an overly idealistic view of the future on one level of the policy process to operate at two levels: the idealistic and the realistic levels. The policymaker would, say, produce idealistic plans to satisfy political principals, but would proceed with his day-to-day work on the basis of more realistic plans, in the full knowledge that idealistic goals would not be attained. If the incentive to produce idealistic plans was stronger than the disincentives associated with failing to reach goals, it would be completely rational for the policymaker to contribute towards the setting of unrealistic targets. The scenario painted here of the rational policymaker working on two levels is perhaps optimal, within a given political context. However, what one imagines Psacharopoulos (1996) was complaining about, was a scenario in which policymakers fail to make the distinction between the realistic and the idealistic, between analysis and politics, and behave according to a set of criteria which are unsustainable and dishonest. The problem with such a scenario would be that realistic work would receive no attention as it would be work considered unworthy of attention, because it fell so far short of the ideal. Pritchett, Woolcock and Andrews (2012) offer a compelling account of the problem being described here, in an article titled ‘Looking like a state: Techniques of persistent failure in state capability for implementation’. Their term ‘isomorphic mimicry’ is more or less a synonym of the term ‘ritualism’ used by Inbar (1996). Pritchett, Woolcock and Andrews (2012) explain how the problem is often compounded by advice given to developing countries by reputable policy experts, who inadvertently undermine institutions by encouraging the over-stretching of existing and very limited institutional capacity, the result being ‘mimicry’, or a world of pretence removed from a sense of what can be realistically done.

The discussion would be incomplete without some explicit reference to ideology, given how frequently the problems described above are expressed in terms of the undue influence of ideology on policymaking. The Oxford Dictionary defines ideology as ‘a system of ideas and ideals, especially one which forms the basis of economic or political theory and policy’[4]. Downs (1957) presents an economics-focussed description of ideology, which deals with a few of the more value-laden aspect of the term, specifically what one could refer to as ‘good’ and ‘bad’ ideology. In a democracy, ideology becomes a tool for organising voter choice in a context of information asymmetries. People are unable to understand the full range of policy choices made by specific parties, and hence parties construct ideologies as shorthand descriptions of what voters can expect from any party should it become or remain the governing party. Ideology is thus a necessary device to deal with policy complexity. Ideology becomes ‘bad’, or malign, when rent-seeking politicians pay too much attention to maintaining and adjusting the ideology in the interests of attracting votes, as opposed to paying attention to actual policies. It is not difficult to imagine how this may result in the ritualism already referred, in particular if the maintenance of ideology through continuous production of highly idealistic policies, combined with tactics to detract from real issues such as the persistence of poverty, become pervasive not just amongst politicians, but most senior government bureaucrats.

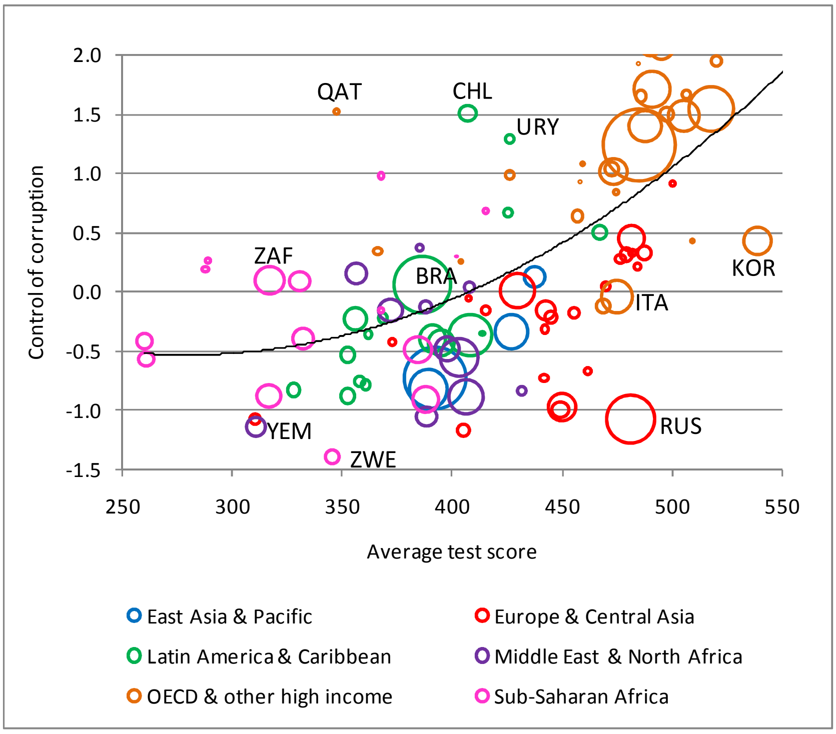
[4] From http://www.oxforddictionaries.com.

Ideology, including 'bad' ideology that detracts from facts and how the population in general experiences the world, is pervasive and not just present in the minds of politicians and bureaucrats. Economists of course need ideology as a system to organise their ideas and ideals. But they sometimes allow ideology to take over and cloud the facts. There is lively debate within economics about this problem. The arguments made by Pritchett *et al* (2012) are partly a rebuke aimed at fellow economists guilty of promoting overly idealised, and thus impractical, notions of the kind of state that can be realised within developing countries. Ideology has blinded these economic advisors to the realities of capacity constraints and institutional factors working against greater government effectiveness. But an excess of ideology not only detracts from fact, ideology also has a tendency to create the impression of large differences in policy standpoints when actually these are rather small. In the language of Downs (1957), political parties in democracies need to create sufficient distinctiveness in their ideological 'brand', even if their policy positions may not be that different from each other. Differences across the policy bundles of different parties are thus emphasised, and similarities are downplayed. It will be argued below that the language of ideology has created an impression of a deep rift between certain educationists and certain economists, which to a large degree disappears when one examines the policy detail. This is not to deny that differences in policy positions exist, or that people who have studied economics tend to adopt rather specific sets of policy values, values which would tend to be considered on the right of the political spectrum, for instance a preference for fewer trade restrictions (Allgood, Bosshardt, Van der Klaauw and Watts, 2012). However, if there is a fundamental ideological difference between the educationists and economists in the apparent stand-off that will be discussed below, it is arguably that economists, with their rather strong insistence that policy advice must be based on empirical evidence, combined with their interest in data analysis techniques, tend towards an empiricist conservatism, and a rejection of romanticism. Economics is the dismal science. Economists, with their attachment, through data, to empirical realities of the past and present, combined with their theoretical emphasis on constraints, are not inclined to believe that human behaviour will be fundamentally different in the future to what it is now. This stands in stark contrast to the emphasis in the field of education, which is not just on educating people, but often also on changing human behaviour in fundamental ways.

Returning to Psacharopoulos's complaint about the amount of politics, as opposed to analysis, in education, arguments made in Part III of the dissertation will partly be based on the assumption that economics in general, and in particular good economics, is less clouded by ideology, and more practically oriented toward problem-solving, than much of the literature in the field of education. Obviously there is some very practical advice emerging from the education specialists, and economists are sometimes guilty of giving impractical advice. Yet some of the focus in Part III of the dissertation is on building a case for more economic thinking within education policymaking, not just because the opportunities for this seem promising, but also because the use of economic analysis and thinking to strengthen education policymaking in the past have borne fruit.

Corruption and, related to this, a slack approach to policymaking, can result in a situation where policies, insofar as they exist at all, do not guide practice and practice is instead guided by informal, and often inequitable, conventions. Figure 63 illustrates the correlation between corruption and level of educational quality, using for the former values from the World Bank's World Governance Indicators, which agree to a high degree with the corruption perceptions index of Transparency International (2011) (the correlation coefficient is 0.985). The graph suggests that to some degree less corruption is associated with better educational quality in the development process, though the graph also points to notable exceptions, such as Russia (high corruption and high educational quality) and Qatar (low corruption but also low educational outcomes). China, which is not included in the graph because average test score data are not available, carries a value of -0.6 on the vertical axis, in other words it is viewed as more corrupt than either South Africa or Brazil. One might expect greater difficulties in identifying what policies guide governments in more corrupt countries, given that corruption by its nature requires a high degree of informality and deviation from rules and laws.

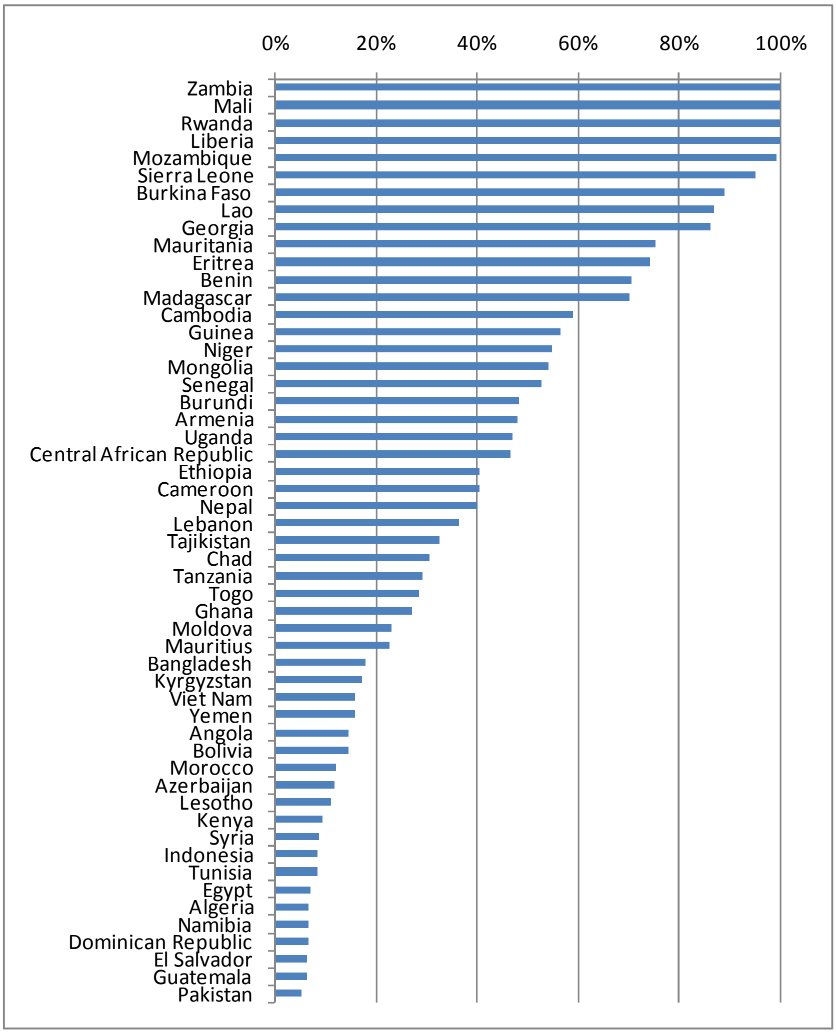
Figure 63: Corruption and quality of education



Sources: World Bank, 2010 (control of corruption); United Nations Statistics Division, 2009 (population). See section Part I of the dissertation for details on how average test scores were calculated.
Note: The area of each bubble reflects population size. The trendline reflects the quadratic function, with adjusted R^2 being 0.355.

A further complicating factor when policies are identified in the case of many developing countries is the strong presence of aid donors. Such a situation may result in a fragmented policy landscape in a country where the unifying role of the state is undermined by multiple policy agendas of several donor countries. The following graph underlines how widespread the presence of donors is in Africa. Of the 33 countries where 20% or more of the education budget is funded by donors, 24 are in Africa.

Figure 64: Percentage of education budget that is donor-funded



Source: UNESCO, 2011a (variables used were total population, GNP per capita, public expenditure on education as a percentage of GNP, and total aid to education).
Note: The horizontal axis refers to overseas development aid received for education divided by public education spending, all in 2008. Discrepancies between values results in the percentage exceeding 100% for certain countries. Countries represented are those with a population of over 1.0 million and with a percentage exceeding 1%.

Poor document availability and the multi-layered nature of education policy in developing countries discussed above are obstacles that make the identification of policy priorities in developing countries difficult. The methodology implicit in the OECD's series of reviews of national policies for education provides some idea of what methodologies are possible and informative. A few of these reviews deal with non-OECD countries, notably reviews for South Africa, Chile, China, Dominican Republic and Russia have been produced[5]. These reviews suggest that both the more idealistic policies and the policies, explicit or implicit, that actually govern implementation should be analysed. The former, though possibly unrealistic, reflect important signals that are being sent through the education system and could point to the direction that will be taken in future policy reforms. The latter, on the other hand,

[5] See http://www.oecd.org/edu/reviews/nationalpolicies.

are important for understanding what concrete choices a country has made within its resourcing and political constraints. Law and Pan (2009: 228) put forward the terms 'law of book' and 'law in action' to describe the two levels of policy. It seems convenient to think of the second layer of policy, the 'law in action', as consisting of two elements. On the one hand, the systems, which could be partly informal, that allocate public and private funding within the education sector are a concrete expression of a country's policy priorities. On the other hand, the presence or absence of incentives (and disincentives) built into the education system to organise principal-agent dynamics, for instance rewards (monetary or non-monetary) recognising effective management, and sanctions against poorly performing teachers, represents a further concrete expression of policy choices made.

The specific policy positions of the economists will be discussed in this section in terms of external and internal efficiencies. This framework will also inform the analysis of policy positions when the four developing countries are discussed. This latter discussion is partly aimed at demonstrating that the economic framework used is a practical and informative one if one wants to identify key elements of what are ultimately highly complex systems. One could of course delve much deeper into the policy positions of the economists than what is done below, but the aim is to offer a discussion that is sufficient to create a framework for the discussion of actual country-specific policies.

With respect to external efficiency, or the degree to which the education system contributes towards individual income improvements and national development, the World Bank's (2011) education strategy document emphasises the need for a stronger focus on qualitative improvement at all levels of a country's education system. The importance of quality improvements at the basic education level receives special emphasis. Using the kinds of statistics dealt with extensively in this dissertation, the magnitude of the effect is explained within the World Bank document as an additional two percentage points in annual GDP growth resulting from a one standard deviation improvement in learning outcomes in schools, where the reference is to a standard deviation across individual pupils in an international testing programme such as PISA[6]. But the World Bank (2011) also places emphasis on improving the quality of education at the post-school level, particularly in tertiary education. Education's role in development is not seen as limited to improvements in national income. It also contributes to better employment rates, including more self-employment, and improved human capital in health terms. Education as an equaliser in terms of socio-economic status and gender disparities is emphasised not only from a justice perspective, but also because, it is argued, more equity is associated with better mean values in the development indicators. For instance, societies with a more equitable distribution of income tend to be societies with higher overall levels of income. In education, higher and more equitable participation rates in post-school education are seen as an increasingly important policy priority for developing countries.

The 2010 *Human development report* of the UNDP (2010), sub-titled *The real wealth of nations: Pathways to human development*, provides an interesting indication that the appreciation of the centrality of educational quality is not always present in the way advocated by most of the recent economic literature, such as the World Bank's (2011) strategy and Hanushek and Woessman (2009). The 2010 UNDP report in fact

[6] Programme for International Student Assessment.

focuses mainly on levels of participation in schooling, with rather little focus on the quality of schooling received, in a manner that is reminiscent of the economic literature prior to around 2000. Clearly, an adequate focus on the quality of education can still not be taken for granted, even in the policy statements of bodies of the UN.

Two key external efficiency questions that should be important to education policymakers seem to be insufficiently explored in economically-focussed policy statements such as that of the World Bank (2011). One is the degree to which streaming, or the special nurturing of more talented individuals through interventions such as schools for the gifted, are important for national development. As will be seen in the discussion on China, clearer guidance from economists in this regard seems needed. Some advice is provided by Hanushek and Woessman (2009), who, in considering the trade-off between 'rocket scientists or education for all' conclude on the basis of their data analysis that both excellence at the top end of the range of educational outcomes, and better average results, are separately important for economic growth. The second question relates to vocational education. What kind of balance between more academic post-school education and more vocationally oriented education should policymakers maintain? This appears to be a question that has been under-analysed by economic think tanks such as the World Bank.

The tension between politics and analysis in education is evident in some of the critique directed at economic interpretations of education. With respect to what the economists would refer to as external efficiency issues, the oppositional literature tends to reject the notion that the education system should be designed in a way that best serves the broader economy and the labour market. There are two rather different arguments behind this opposition, a professional one and a political economy one. The professional argument is, roughly, that trying to align the education system to economic needs tends to result in crude educational solutions that ultimately fail because they do not appreciate the many facets, including non-economic ones, of any education system. Very crudely, the argument runs that the economists do not know enough about education to tell the educationists how to do their job. The political economy argument, again somewhat roughly, sees the capitalist economy as unjust and considers that the education system offers a key opportunity to change this economy through the creation of a post- or anti-capitalist generation. On ideological grounds then, instead of getting the education system to maintain and grow the existing economy, the education system should be used to bring about fundamental structural change in that economy. Delandshere and Petrosky (2004: 2), who generally support this argument, describe education reforms as conservative attempts aimed at preventing questioning of the economic status quo:

> Workers with critical thinking and problem-solving skills might not easily adapt to the economic requirements of business. In the face of uncertainty, contradictions, and the unknown, reformers have rallied around predefined standards, strict policies, and strategies, so that we have the illusion of knowing where we are, where we are going, and what and how we should teach.

Where Delandshere and Petrosky (2004) seem correct is in assuming that government reformers of the education sector in the United States are 'pro-business' in the sense that they have no intention of changing the basic capitalist system. But these ideological matters should not cloud the key fact that the existing education system has tended to fail in its attempts to generate problem-solving skills amongst sufficient numbers of students, in particular socio-economically disadvantaged students. Anti-reform arguments of the kind offered by Delandshere and Petrosky (2004) often ignore the evidence on past successes and failures, and consider the *intentions* of teachers to be a sufficient sign of success. What is also absent in Delandshere and Petrosky (2004) is an acknowledgement that the reformers in fact do prioritise more problem-solving, partly because business in fact requires more, not less, of this skill.

Turning to internal efficiency issues, or how well education institutions (supported by households) produce desired educational outcomes, the World Bank's (2011) strategy emphasises using a variety of measures affecting human development, starting from the pre-natal stage, in a manner that promotes health and optimal cognitive development. Child nutrition and quality pre-schooling are thus important policy priorities. At the primary and secondary schooling levels, improving the quality of learning outcomes through, in particular, more effective teacher training and educational assessment systems is seen as key. Bruns, Filmer and Patrinos (2011) stress that in developing countries a back-to-basics approach that includes practical accountability systems linking schools to the administration and schools to communities should be promoted. Robust accountability systems should be accompanied by sufficient freedom for schools and teachers to organise their work in ways that respond to local needs whilst they promote system-wide learning goals. In more economic terms, principal-agent relationships between, on the one hand parents and teachers and on the other between the administration and teachers, should be improved through the resolution of existing information asymmetries whereby, for instance, parents do not understand whether their children are in fact learning to read as they should. Better accountability systems are also needed at the tertiary education level, according to the World Bank (2011) strategy. Moreover, in the developing country context there need to be enough second-chance opportunities for youths who fail their first attempts at obtaining a post-school education.

These internal efficiency proposals have also attracted opposition, as explained by Loveless (2005: 25) in the context of the United States:

> Formidable political factors are arrayed against accountability. Progressive education supplies an ideology hostile to test-based accountability. Politically powerful teachers unions represent actors with interests divergent from the aims of accountability systems. Governance in education is both open and porous, providing multiple venues for challenging implementation. Parents, students, and individual teachers (that is, teachers acting alone, not as union members) are the opponents of accountability most prominently featured in the press.

Very significantly, the recent 2011 education policy position paper of Education International (EI), the largest international federation of teacher unions (and according to the EI website[7], the largest international federation of unions across all trades), strongly opposes the policy positions of, for instance, the World Bank.

> The social values of education require public authorities to protect the education sector from the neo-liberal agenda of privatization and commercialisation. This negative agenda includes marketisation and trade in education and intellectual property, the casualisation of employment in the education sector, the application of private-sector management models on education institutions, the privatisation of provision, and the intrusion of for-profit motives or

[7] See http://www.ei-ie.org.

> business interest in the governance of education institutions. (Education International, 2011: 2)

Here it appears as if ideological differences relating to the kind of society one wants at some point in the future, specifically whether a more socialist economy is preferable, cloud the more immediate and practical issues, and make it appear as if there are stark divisions with respect to the latter, when these divisions are perhaps not so large. In at least two senses, the EI and the 'neo-liberal' reformers are closer to each other than what the above extract would suggest. Firstly, both advocate a change towards a more equal society, though the EI sees this change coming about more through changes in the patterns of physical capital ownership, whilst the reformers have emphasised a more equitable distribution of human capital. Secondly, though the reformers may be proponents of 'privatization and commercialisation' in the economy as a whole, the literature that guides them as far as the education sector is concerned acknowledges, on the whole, the difficulty and even the undesirability of bringing about deep structural change, such as privatising entire public schooling systems, and generally focuses on more marginal changes of procedure, such as those brought about my more regular accountability.

The fact that teacher unions from a large range of countries, both developed and developing, are members of EI, added to the fact that teacher unions tend to exert strong pressure on policymakers, makes the EI position particularly noteworthy. (The most noteworthy gap in the IE members list is the absence of any teacher union from the People's Republic of China. In the case of South Africa three different unions are members.) Perhaps the most serious area of possible contention relates to assessments aimed at improving the accountability of education institutions. Such assessments are strongly promoted by the economic literature as a means of tackling the considerable information asymmetries inherent in education systems, but at the same time face stiff opposition. The EI's 2011 policy statement argues:

> EI believes that a widespread abuse of the notion of quality to justify standardised forms of testing is harmful to the education system as a whole, as it attempts to reduce the teaching and learning process to quantifiable indicators. It is the standardization and one-dimensional approach to testing and evaluation of the teaching and learning processes to which EI objects strongly. ... Public authorities should guard against the potential misuse of the Programme for International Student Assessment (PISA) in the administration and planning of education systems. (Education International, 2011: 4)

Yet in another part of the EI's statement, a more conciliatory stance is taken. Here there is less of an outright rejection of standardised assessments, and more an implicit acceptance that if used correctly, standardised assessments have a role to play.

> When one form of evaluation designed for a particular purpose is used to serve a different purpose, the consequences can be unforeseen and damaging. (Education International, 2011: 4)

The thrust, however, of the EI's statement is that standardised testing is bad for education, and, implicitly, for the struggle for a better and more equal society. Reference can be made to three arguments that education economists would commonly make, which render the EI's arguments largely invalid. The first is the argument, supported by much evidence, discussed in for instance Bruns, Filmer and Patrinos (2011), pointing to the fact that education systems have been notoriously bad at educating the poor, in fact much worse than is commonly believed. The second argument is that evidence points rather clearly towards inequalities in human capital being the fundamental underpinning of other inequalities, in particular income inequality. What the EI is thus ignoring is that much of the social change they would like to see must be achieved through better educational outcomes for the poor. Thirdly, the EI ignores the empirical evidence suggesting that the accountability reforms it rejects have in fact succeeded in improving education for the poor.

The methodological debates surrounding the kinds of impact evaluations that inform the policy conclusions of Bruns, Filmer and Patrinos (2011) are so important they need to be taken into account in any examination of education policy. A country's education policies are not just about what actions to take, but also how feedback and research are used to inform and fine-tune actions. Formal impact evaluation techniques, using treatment and control groups, along the lines of research in the medical profession, offer a compelling ideal for how to generate the knowledge needed for policy reform and action. However, any practical national strategy on policy evaluation would have to acknowledge that the costs and logistical and political complexities associated with formal impact evaluations such as randomised control trials are substantial and that at best one can subject a small minority of policies and interventions to formal impact evaluations. Added to this, is the fact that certain interventions are not amenable to formal impact evaluations in the first place. These problems are acknowledged by many impact evaluation proponents, such as Duflo, Glennester and Kremer (2006), who warn, for instance, that impact evaluations are not appropriate where there are large expected general equilibrium effects resulting from the expansion of an intervention from a project to a system-wide programme. For instance, a small pilot project experimenting with the impact of monetary incentives for teachers might render positive results, but the benefits may be outweighed by social costs if the intervention is taken to scale and teacher unions object to the incentives on ideological grounds. The literature suggests that the good policymaker needs to do two things. On the one hand, formal impact evaluations need to be employed with respect to carefully selected interventions where evaluation results are likely to be relevant and externally valid, meaning applicable on a larger scale. On the other, the policymaker needs to ensure that there are enough 'second-best' and practical evaluation activities, even if they do not reach the 'gold standard' of, say, the randomised control trial (Schlotter, Schwerdt and Woessman, 2010: 6). The literature on 'second-best' approaches is not large, though there appears to be an increasing realisation that this is a gap that needs to be filled. An example of an interesting contribution would be Pritchett, Samji and Hammer (2013), who acknowledge institutional and incentive problems around the implementation of fully-fledged and formal impact evaluations, and advocate novel approaches that are more sensitive to the way policymakers actually reform systems, and focus on what can be learnt from variation within interventions.

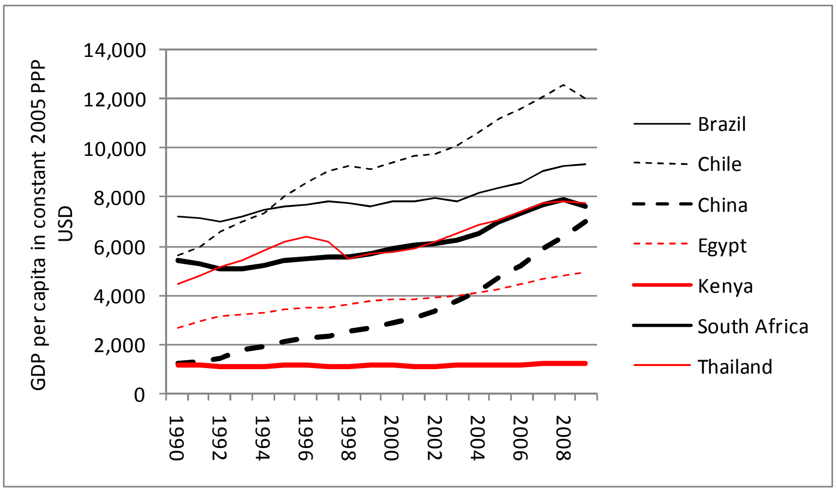
14 POLICY PRIORITIES IN FOUR DEVELOPING COUNTRIES

14.1 Introduction

What policy positions have developing countries governments adopted, relative to the recommendations of the economic literature but also the opponents to these recommendations? This question is answered in section 14 with reference to a group of four important but also diverse countries. South Africa is included because South Africa is dealt with elsewhere in the dissertation. Brazil was selected as it has increasingly been considered a best practices model for poverty alleviation within a non-centralised and democratic governance structure. Chile has often been considered an innovator amongst developing countries in the area of education policies. China was selected due to its rising geo-political importance and because favourable human capital conditions are believed to lie behind its rapid economic growth since the 1980s.

The following graph provides some economic development context to the discussion. Of the four countries, Chile and China have experienced particularly strong growth in GDP per capita. Such patterns should influence these countries' education policy priorities. One might expect Chile and China to pay more attention to the question of how to allocate additional public funding towards substantial growth in the education sector. In Brazil and South Africa, on the other hand, one might expect a stronger emphasis on educational improvement achieved through efficiency gains. Brazil and South Africa should also pay close attention to the degree to which human capital constraints have impeded faster growth and attempt to remedy this situation, partly by learning from the experiences of countries such as Chile and China which have experienced rapid growth.

Figure 65: GDP per capita trend in key developing countries



Source: Heston, Summers and Aten, 2011.

In part, the discussion and comparison of the four countries is intended to demonstrate an effective way of undertaking a comparison of different education systems. The complaint made by Psacharopoulos (1990) about the unsystemic and unquantitative nature of many articles published in arguably the world's foremost journal dealing with comparative education studies, namely the *Comparative Education Review*, is a good reference point. A short meta-analysis of articles from the journal from the years 1981 (a year within the time range that Psacharopoulos's was presumably thinking of) and 2012 was undertaken to explore the dimensions of the alleged problem with the journal. Results, based on an attempt to answer a few basic questions about each article, appear in Table 40 below. Psacharopoulos's complaint is at least partly supported. In 1981, no articles includes all the features looked for: an actual comparison, some discussion of specific policies, statistics, and some new analysis of microdata. In 2012, 19% of articles had all these features. However, the majority of articles in 1981 had at least some statistics, even if they were drawn from other sources, so the accusation of insufficient quantification is not strongly supported. What seems more striking is that fewer than half of the 1981 articles included an actual comparison. By 2012 this had improved to 75%.

Table 40: Patterns in *Comparative Education Review* articles

	Is there a comparison between different education systems?	Are the effects or problems of specific policies discussed?	Are statistics that provide a sense of magnitudes included?	Are some of the statistics derived from new analysis of microdata?	% of 2012 articles with this comb-ination	% of 1981 articles with this comb-ination
Combinations of 'yes' responses (just 3 most common combinations from 2012 shown)						
Comb. A	1	0	1	1	25	10
Comb. B	1	1	1	1	19	0
Comb. C	1	0	0	0	13	10
% of articles with response 'yes'						
2012	75	44	81	56		
1981	45	50	70	15		

Source: Own analysis of Comparative Education Review.
Note: For each of the two years, all articles from three quarters which did not have quarterly-specific special topics were examined. In total, 16 articles from 2012 and 20 from 1981 were examined.

This brief *Comparative Education Review* analysis suggests that there have been improvements over time in comparative education methods. However, none of the articles examined has what is presented below, namely a policy comparison of a few countries within an economics of education framework. The benefits of such a framework are discussed, in particular in the conclusion of section 15. It should be added that the piecing together of elements of even one country's education system, with no intention of any cross-country comparison, in order to answer over-arching questions around internal and external efficiency, and equity, is surprisingly rare. In particular, the description of China's education system included below can be considered a modest contribution towards the filling of an important gap, namely the absence of analytical and critical descriptions of China's education system in English.

14.2 South Africa

The policy position on education's role in the economy has become clearer in South Africa over time. The 1995 *White paper on education and training*, which was seen as a basic blueprint for post-apartheid education, referred in general terms to the role played by education in preparing youth for economic participation (South Africa: Department of Education, 1995). Educational quality is referred to in a way that is likely to result in multiple interpretations of the concept. One could be dealing with a disciplined system, a well-resourced one, one with well-qualified teachers or, as the economists would now emphasise, a system where there is sufficient learning amongst pupils. More recent policy documents are much clearer and are far more aligned to the learning outcomes definition. The minister of basic education's 'delivery agreement' includes the following in its introduction (South Africa: Department of Basic Education, 2010a: 4):

> ...the President referred to the vital role of the education system in improving productivity and competitiveness in the economy. The President also underlined that 'our education targets are simple but critical'. The need is fairly straightforward as far as the basic education sector is concerned. Our children and youths need to be better prepared by their schools to read, write, think critically and solve numerical problems. These skills are the foundations on which further studies, job satisfaction, productivity and meaningful citizenship are based.

One key national policy document expresses the educational targets in terms of standardised test score improvements: the average performance of mathematics pupils in grade 8 should increase by 140 TIMSS[8] points in 15 years, meaning an improvement of 1.4 standard deviations across TIMSS pupils (South Africa: Department of Basic Education, 2010b: 51). The minister's delivery agreement argues that increasing per pupil spending is not a high priority, mentioning that this spending is considerably higher than in the comparable Latin America region. The need to reach a 100% participation rate in pre-school in the year immediately before the first grade is emphasised. The current rate (in 2010) is said to be around 80%. A national development plan produced by the country's national planning commission goes further and prioritises full participation for two years of pre-school (South Africa: National Planning Commission, 2011: 270, 279).

The delivery agreement considers failings and gaps in the accountability and incentives mechanisms of the schooling system an important obstacle to improvements in learning outcomes. Here the policy emphasis for the immediate future falls on strengthening a standardised assessment system introduced in 2008. This system is intended to reduce the over-reliance on the national examinations at the end of upper secondary schooling as means of holding schools accountable for the quality of their services. The need for strengthening both accountability routes, one to parents and the other to the administration, partly through the new assessments system, is acknowledged. Other priorities include a proposed system whereby teachers would be more accountable for the use of their time, specifically their pacing of sections of the national curriculum across the school year. Some reverse accountability is proposed in the form of ratings by schools of the quality of support received by the administration.

The delivery agreement acknowledges the importance of pre-service and in-service training for teachers, but how this should be strengthened, for instance with respect to optimal in-service training methodologies and the selection of youths for pre-service training, is insufficiently clear. A separate teacher development plan published in 2011 provides some additional details and a few insights into how economics of education could play a greater role in informing teacher development, not just in South Africa (South Africa: Department of Basic Education, 2011a). The plan does acknowledge the importance of providing, to a greater degree than in the past, in-service training that responds to demand and real needs amongst teachers. It also implicitly acknowledges the need to consider cost through an emphasis on distance education. However, the plan also displays gaps that are commonly found in

[8] Trends in International Mathematics and Science Study.

education policies. Above all, the importance of monitoring impact is not acknowledged. In the case of teacher training impact could be measured in terms of, for instance, changes in teacher knowledge, teacher behaviour on the job, or the learning outcomes of pupils. An insufficient focus on understanding the impact of costly interventions is also evident in a 2011 policy on standards that should be applied in teacher pre-service training. The weak emphasis on the monitoring of impact seems to be less an unintended omission than a conscious scepticism towards the measurement of outcomes, as suggested by the following extract from the 2011 standards document (South Africa: Department of Higher Education and Training, 2011: 7):

> A purely skills-based approach, which relies almost exclusively on evidence of demonstrable outcomes as measures of success, without paying attention as to how knowledge must underpin these skills for them to impact effectively on learning, produces technicists who may be able to replicate performance in similar contexts, but who are severely challenged when the context changes.

Teachers need to be trained in a manner than assists their adaptability to different contexts, but why measurement of outcomes should be in conflict with this principle is not clear. Better measurement of outcomes in teacher training within the tertiary education system in fact seems necessary given the suspected degree of inequality between institutions discussed in a 2011 green paper on post-school education (South Africa: Department of Higher Education and Training, 2012). A further indication that the benefits of monitoring quality in teacher training are not fully appreciated can be found in a comprehensive 2010 review, by the government, of teacher training programmes. This review evaluates programmes in terms of their intent and design in an attempt to compare training institutions, but includes virtually no consideration of outcomes, something that could have been partially dealt with through, for instance, an examination of the written work of teacher trainees (South Africa: Council on Higher Education, 2010). Even better, however, though more costly, would have been a consideration of more downstream outcomes, implemented for instance through interviews with graduates already working as teachers and observations of classroom practices.

The national development plan pays attention to a matter not often dealt with explicitly in education plans, namely corruption. Specifically, nepotism in the appointment of teachers and school principals is identified as a problem. One solution put forward by the plan to deal with this is competency tests for anyone applying for a school principal position (South Africa, 2011: 270, 281).

The green paper on post-school education emphasises enrolment growth, in particular as far as socio-economically disadvantaged students are concerned, and quality enhancement in the sector. However, the magnitude of the enrolment growth envisaged seems to be informed more by political aspirations than empirical analysis. There is no benchmarking of the envisaged enrolment trend against those in other countries, or consideration of cost implications in terms of, for instance, total spending over GDP. More or less, the green paper envisages a situation in South Africa in 2030 where post-school enrolments would be around 4.7 times as large as a single youth age cohort, a level twice that currently found in Chile, a developing country with particularly high levels of post-school enrolments. However, the envisaged South African level would still be half of the level found in the United

States (Gustafsson, 2011: 18[9]). The need for growth in the post-school sector is clearly needed as suggested by the skills shortage in the labour market. Gustafsson (2011: 14), using household data for eight developing countries, shows that whilst attainment of up to twelve years of education is exceptionally high for South Africa, attainment of thirteen or more years of education is exceptionally low. Relatively speaking, there has been over-investment in enrolments in schooling and under-investment in enrolments in post-school institutions. There is a strong emphasis on increasing the number of PhDs in critical areas within the green paper. For instance, in engineering a target number of PhDs in 2030 that is five times the 2011 level is put forward. The need to produce enough lecturers that can sustain the envisaged growth is also emphasised, though specific strategies in this regard are not explained and the important matter of whether growth in lecturers needs to be proportional to the growth in students is not discussed.

The need for improving the quality of learning outcomes in universities and other post-school institutions is recognised. That this partly depends on better learning outcomes in schools is mentioned, but not sufficiently explored. The green paper argues that the problem is not that official standards set by government and universities are not high enough. Instead, the problem is that quality assurance mechanisms and institutions are too complex and insufficiently developed. The exact nature of this critical problem and proposed solutions are not clear in the green paper, however. What is made clear is that universities should continue to quality assure themselves whilst other post-school institutions should continue to be subject to centralised quality control mechanisms, such as national examinations. A key question that is implied is how university autonomy with respect to quality assurance is compatible with qualitative improvements in the future, if the existing quality assurance mechanisms for universities have clearly not functioned well in the past, in particular as far as academically weaker universities are concerned.

Do the policy positions of South African policymakers concur with those found in the economic literature? At a strategic level, there appears to be concurrence, especially as far as the primary and secondary education policies are concerned. At these levels, the importance of learning outcomes and accountability of institutions to the administration and parents, is emphasised. At the tertiary level, and in the policies on teacher training, there appears to be a lesser awareness of having the right accountability structures in place. At all levels, what appears to be lacking is operational depth in the form of detailed operational reports, impact evaluations or sufficiently clear medium-term plans. This is perhaps not surprising given the newness of many of the strategic documents. Achieving sufficient operational depth should be a priority for the education authorities.

14.3 Brazil

To gain an idea of how education policy thinking has evolved in Brazil in recent years, it is informative to compare the national education plan of 2001 to a proposed national plan released in 2010 (Brazil: Ministério da Educação, 2001, 2010). The 2001 plan focussed largely on increasing enrolment ratios and hours per day spent at school. In the case of Brazil the latter priority is closely linked to a desire to reduce double and triple shifting in schools, which remains high. For instance, in 2010

[9] Also Current Population Survey 2010 spreadsheets at http://www.census.gov.

around 41% of pupils at the Grades 1 to 8 levels were attending a shift which was not the morning shift, generally the afternoon shift (the third shift is the evening shift) [10]. By the time of the 2010 plan, however, the emphasis had largely shifted from enrolments to learning outcomes, partly because the previous decade had seen major advances in the enrolment ratios.

The 2010 plan is specific about the monitoring of educational quality in schools. At the centre of the monitoring process is an index, the Index of Basic Education Development, or IDEB[11]. This index value combines a measure of pupil performance with a measure of grade progression. The latter is included to counteract perverse selection effects arising out of the focus on pupil performance, for instance in the form of the pushing out or excessive grade repetition of academically weaker pupils. IDEB values are calculated according to the following formula:

$$IDEB = N^{\alpha} P^{\beta} \qquad (89)$$

N is the average score for a grade (for instance nationally or in a single school) and P is an indicator of grade attainment based on the grade promotion rates of the grades in the curriculum phase leading up to the tested grade. α and δ are weights attached to the two factors, whose sum is 1.0. P, the indicator of grade attainment, is calculated as follows:

$$P = \sum_{r=1}^{n} \frac{n}{p_r} \qquad (90)$$

where p is the promotion rate for a specific grade, in other words the percentage of pupils who do not drop out or repeat, and n refers to the number of grades within the phase (Fernandes, 2007). IDEB values from the national to the school level are published on the federal ministry of education website[12]. There is a downloadable report for each school where the school's results are compared to those of the municipality as a whole, the state and the country. The pupil performance values in IDEB draw from Prova Brasil, a programme introduced in 2005 which involves the testing of pupils from all schools in selected grades. Prova Brasil emerged out of a sample-based national assessment system that evolved over more than a decade before 2005. The IDEB, introduced in 2007, has come to be used by a number of states and municipalities as a basis for monetary incentives that reward schools whose performance improves (Parandekar, Amorim and Welsh, 2008; Bruns, Filmer and Patrinos, 2011). More generally, however, economists such as the World Bank's Barbara Bruns have attributed Brazil's performance improvements over the years in PISA to better accountability and the counteracting of information asymmetries brought about by Prova Brasil, even without the introduction of monetary incentives (Bruns, 2010). (PISA assesses pupils aged 15 regardless of the kind of education institution in which they are enrolled.) A similar interpretation is found in an OECD (2012) report on Brazil's educational improvements. The 2010 plan explains that the pupil performance component of IDEB is benchmarked against PISA results and envisages a PISA improvement of 80 points, or 0.8 of a standard deviation, by 2020.

---

[10] Spreadsheets in the *Censo da Educação Básica* series at http://portal.inep.gov.br/basica-censo-escolar-sinopse-sinopse.
[11] Acronym for Índice de Desenvolvimento da Educação Básica.
[12] See http://provabrasil.inep.gov.br.

A key enrolment target in the 2010 plan is to universalise participation for the population aged 4 to 17 by 2020. Moreover the aim is for 25% of enrolment at the secondary level to be vocational. There is considerable emphasis on recruiting more teachers and improving the existing workforce through in-service training. Brazil is one of very few developing countries where the academic literature points to success in teacher in-service training (Bof, 2004). Specifically, Brazil is said to have achieved a degree of success in reaching a large number of teachers with a mix of distance and contact training, characterised by regular testing of trainees and monitoring of processes such as participation. Moreover, Brazil's in-service training efforts are noteworthy for their success in bringing together different public and private training organisations. The 2010 plan envisages better pay for teachers, free pre-service training for those wishing to become schoolteachers and, in an attempt to improve the quality of teaching, an entrance examination for those wishing to enter the profession. As in South Africa, a special entrance examination for those wishing to become school principals is envisaged.

At the post-school level, substantial enrolment growth is planned. The gross enrolment ratio for the population aged 18 to 24 is expected to increase to 50% by 2020 according to the 2010 plan. This is probably unrealistic if one considers that the figure for the United States in 2010 was 47%[13]. Enrolment growth is expected to be possible due partly to more extensive state guaranteeing of student loans. To promote quality in universities Brazil employs standardised testing of all university graduates. Tests are discipline-specific, thus for instance economics students write a common economics examination, apart from examinations that the particular university may set. Results from the standardised tests are used, amongst other things, to assess whether private universities should continue to operate (Rezende, 2010). Average scores per institution are publicly available and individual students are given a report with their scores. These reports are commonly used by students to communicate their competencies to potential employers. This intensity of institutional accountability at the tertiary level is unusual amongst both developing and developed countries. The number of Masters and PhD graduates per year is expected to reach 60,000 and 25,000 by 2020. This translates into around 1.7% and 0.7% of the age 25 cohort (United Nations, 2009). The corresponding figures in the United States are 8.5% and 1.6%. Here Brazil's goals appear to be more realistic.

Brazil is in a number of ways where South Africa's policies say South Africa would like to be. It could be argued that Brazil is around ten years ahead of South Africa, assuming South Africa's plans are implemented in the near future. Brazil has around ten more years of experience than South Africa in terms of national testing systems. Prova Brasil is similar to what South Africa would like to realise through its national assessment system. Brazil has had greater success than South Africa in rolling out large-scale in-service training for teachers.

14.4 Chile

What is striking about Chile's education policies and systems is how clear the policies are, how deeply rooted a tradition of accountability (of the type World Bank economists would espouse) is, and the relatively large volumes of research available on Chile's education strategies. The greater availability of research and

[13] Current Population Survey 2010 spreadsheets at http://www.census.gov.

documentation (in English or the national language) in the case of Chile relative to Brazil is particularly noteworthy. A manifestation of Chile's commitment to democracy and evidence-driven policymaking (not just in education) was the country's acceptance into the OECD in 2010, making Chile the third non-high income country to join the organisation, after Turkey and Mexico. Crouch (2005) has argued that in education Chile's success could be attributable to the exceptionally strong presence of economists and economically-minded leaders in the sector, as well as a mature and constructive relationship between the government and the largest teacher union.

McEwan, Urquiola and Vegas (2008) argue that Chile's strong emphasis on educational innovation has not paid off in the form of better learning outcomes. However, they make use of international testing programmes stretching just to 1999. As pointed out by Gustafsson (2012), if results up to 2009 are taken into account, Chile, together with Brazil and Tunisia, are exceptionally good improvers amongst developing countries with respect to the international testing programmes (but Chile's level was considerably above that of the other two countries in 2009). What this suggests is an impact lag which has been longer than what McEwan *et al* (2008) may have assumed.

In contrast to Brazil and South Africa, Chile's most recent policies for the education sector emphasise not so much radical change as a continued focus on strategies that are not new. This is clear in, for instance, Chile's 2011 'law on quality and equity' (Chile: Congreso Nacional, 2011[14]).

In the schooling system, policies that explicitly promote educational quality include a funding system that treats public and private schools equally, thus allowing them to compete on an equal footing, and monetary incentives that reward schools with exceptional improvements. Pre-school participation ratios are good even by developed country standards: 62% of Chile's children aged 3 to 5 are enrolled, against an OECD average of 77% (the figures for Canada, Australia and the United States are lower than Chile's) [15]. The school funding system, often referred to as a 'voucher system', has been the subject of much analysis. Following West (1996), the term 'voucher system' has been used to describe such a wide variety of funding systems and is cause for so much confusion and ideological polemic, that the term is perhaps best not used. In short, what makes Chile's funding system noteworthy and unusual is that exactly the same funding rules are applied to all schools, whether public or private, meaning that, at least in urban areas, public schools which under-perform face a real risk of losing all their funding to better performing private schools. The usual protection thus afforded to public schools is largely absent in Chile. The private sector is large: by 2005 around half of all primary school pupils were in private schools (McEwan, Urquiola and Vegas, 2008: 5). Though this funding system may lie behind the progress made in Chile's learning outcomes, it may also have exacerbated social stratification across schools (McEwan, Urquiola and Vegas, 2008).

[14] A reader-friendly synopsis is found at http://www.reformaescolar.mineduc.cl.
[15] See OECD statistical release at http://www.oecd.org/dataoecd/46/13/37864698.pdf.

Chile's standardised testing system in schools, SIMCE[16], has a long and interesting history. Started in 1988, the specifications of the system were legislated into a strongly binding 'constitutional law' in 1990 one day before the end of military rule (Meckes and Carrasco, 2010: 237). Policymakers have focussed not just on improving the reliability of the results of the system, but also its use as an accountability tool. An illustration of this are the 40-page reports generated for every school with a number of different analyses and comparisons and simpler four-page reports directed at parents. These reports are available on the ministry of education website[17]. The system has been allowed to evolve, with more rigorous accountability processes only being introduced when the measurement rigour and degree of trust in the system permitted this. Taut *el al* (2009) have pointed out the importance of not equating policy intention with reality: they find that only half of school principals fully understood the SIMCE reports and that parent knowledge about the system was lower than official reports would suggest (Taut, Cortés, Sebastian and Preiss, 2009: 135). Despite such shortcomings, SIMCE appears to come close to being a best practice case with respect to such systems in developing countries. Details on the system are far more accessible than details on Prova Brasil, for which more information should be available in the public domain. The national incentive system rewarding schools for good performance in SIMCE is known as SNED[18]. SNED bonuses are paid to around a quarter of schools every second year. One of several factors considered is the size of year-on-year improvements.

Binder and Contreras (2008) find that at least when the programme was introduced, SNED was able to bring about improvements of around 0.3 standard deviations to the average of all schools due to what the analyst's call a tournament effect, or the effect of knowing that one may win the reward. This effect was found using a matching procedure. However, the data were not able to provide evidence on the sustainability of this effect over time. What was not found to be significant was the 'gift-exchange' effect, or the effect of continuing improvements after receiving the reward almost as a type of gratitude response to the reward. Here a regression discontinuity analysis was used. The magnitude of the tournament effect appears sufficiently large to warrant serious interest amongst policymakers, not just in Chile, in SNED-type rewards. Moreover, the analysis techniques put forward have important implications for the types of data that should be collected in these kinds of programmes.

Another incentive programme where Chile appears to have put the economic emphasis on incentives into practice is the Variable Allocation for Individual Performance or AVDI[19] programme. Here teachers are paid a bonus if they pass an examination they enter on a voluntary basis (Bruns, Filmer and Patrinos, 2011: 179; Taut, Santelices and Stecher, 2011). This is an innovative way of dealing with a problem most education systems devote considerable attention to, often with limited success, namely the design of in-service training approaches supplied by the state to teachers. The Chilean approach shifts the problem to the teacher and obviates many complications relating to the matching of supply with demand and need. Teachers

[16] National Evaluation System (original in Spanish is Sistema de Medición de Calidad de la Educación).
[17] See http://www.mineduc.cl.
[18] National System for Performance Evaluation of Subsidized Educational Establishments (original in Spanish is Sistema Nacional de Evaluación del Desempeño de los Establecimientos Educativos Subvencionados).
[19] Asignación Variable por Desempeño Individual.

themselves obtain, perhaps by making a private financial investment, the capacity building they believe they need. In fact, Chile's education policies lack the relatively strong emphasis found in South Africa and Brazil on large-scale teacher in-service training. Implicitly, they support the recommendation made in an IIEP guide on in-service training that an over-reliance on traditional top-down training methods is not a good thing: 'It is important to recognise that not all aspects of teacher professional development can be (or should be) addressed in courses. There are many other models of professional development which support teachers' development on a regular basis in the workplace.' (Villegas-Reimers, 2003: 142)

Post-school participation in education in Chile is particularly high for a developing country and is almost on a par with that of the United States (Gustafsson, 2011: 17-18). An OECD report on Chile's tertiary education system highlights as a remarkable success the five-fold increase in the participation rate in the poorest 40% of society in just one generation (OCDE, 2009: 13). This rapid expansion is partly due to considerable growth with respect to private universities (OCDE, 2009: 81). One recently introduced intervention aimed at improving competition amongst post-school institutions, in an attempt to raise quality, is the publication of information on labour market returns associated with education and training received in different institutions. This rather unusual and innovative approach to tackling information asymmetries, specifically the problem of youths not knowing the implications of the specialisation and institutions they select, involves the publication on the ministry of education 'My Future' website of the average pay and employment rates of graduates of specific institutions (including vocational training institutions) in specific occupations[20]. Other information provided includes the drop out rates of different institutions. The data are obtained from the institutions themselves, which are obliged to conduct tracer studies covering at least the first year that graduates spend in the labour market. The website publishes the methodology that institutions must follow.

The OECD report argues that what is noteworthy is that the rapid expansion in access to post-school education has occurred without any major compromising of quality. A further quality enhancement initiative is a partnership project with the World Bank that ran from 1997 to 2005 and was specifically focussed on promoting quality through the rationalisation of curricula, training of tertiary education staff and new information systems (OCDE, 2009: 191). However, the OECD report also points to problems in Chile's tertiary education system. The largest structural problem is that degree programmes in Chile are spread out over too many years and are too theoretical. On average, students take seven years to complete their first degree. Moreover, despite policies aimed at fostering competition between universities, mechanisms such as a peer reviews tend to amount to little more than a 'checkbox ticking' exercise. By law certain universities enjoy an effective monopoly in certain academic fields. Inefficiencies and relatively low levels of public spending result in a high reliance on private fee income, which limits the degree to which inequities in the participation ratios can be reduced. There is no countrywide system targeting poorer households with loans and bursaries for university studies. One symptom of these problems is that too few students obtain PhDs (OCDE, 2009: 12, 16). These problems are serious ones and suggest that overall Chile has been more successful at tackling structural problems at the basic education level than at the post-school level.

[20] See http://www.mifuturo.cl.

14.5 China

With regard to education's role in society, Chinese policies have been relatively consistent for a quarter of a century. Hu (2012: 86), in an analysis of four important sector policy statements in the 1985 to 2010 period, concludes:

> Worth noting is the fact that one thing never changes in any of the four documents: the values of education are directed toward serving the economy, in other words, ... prominence is given to positioning people as 'trained talent' and using education as an instrument.

The centrality of economic development is thus strong. However, the emphasis on how to optimise education outcomes has shifted from fundamental change and reform initiatives to improvement and maintenance. There has also been an increasing propensity towards self-criticism by the government, within the constraints imposed by China's one-party system of governance. A 2010 policy review process was unprecedented insofar as it involved the collection of public inputs, which took the form of around 14,000 letters to the authorities (Walker and Qian, 2012: 163). Hu (2012) identifies two key turning points in the recent development of Chinese education policy: 1999 saw a strong shift towards improving educational quality in the sense of what children learn, in other words in the sense understood by the World Bank. Thereafter, 2010 saw a re-focus on improving educational equity following a period of deterioration in this area, something considered a threat to political stability as well as a violation of socialist principles.

China's very different history and culture relative to the three countries considered previously results in often rather fundamental differences in the way problems and solutions need to be understood, or are understood by the Chinese government. For instance, what is widely regarded as an over-emphasis on public examinations in the past has led to a situation where initiatives to improve educational quality often focus largely on how to reduce the influence of examinations. The typical emphasis found in other developing countries on improving the work ethic in education institutions and increasing time on task is largely absent in Chinese policies given that time on task is often of a level that would be considered excessive in other countries. According to Jing and Yu (2006: 259), in lower secondary schools pupils typically devote themselves to schoolwork from 6 in the morning to 11 at night, six or seven days a week. This seems to be an exaggeration, yet the fact that a claim such as this is publishable in an academic journal is telling. The levels of effort seen in China's schools are often attributed to the strong policy and cultural emphasis on examinations in the past.

There are three key examinations. The *zhongkao* examination occurs at the end of lower secondary schooling, at the end of the ninth grade. The *huikao* occurs at the end of upper secondary schooling, at the end of the twelfth grade. The *gaokao* is used to determine who may enter what types of post-school education and is generally administered by upper secondary schools at the twelfth grade (Hannum, Xuehui and Hua-Yu, 2011: 272).

The hypothesis that Chinese school pupils perform well, a hypothesis often based on the academic performance of Chinese schooled students attending universities outside China, appears to have been confirmed by the fact that in the 2009 run of PISA the city of Shanghai, the only part of China to participate in PISA, obtained results that exceeded those of all other countries, including the city states of Hong Kong and Singapore, in reading, mathematics and science. The People's Republic of China has never participated as a country in an international assessment programme and there is no Chinese national assessment system along the lines of Brazil's Prova Brasil. As a result, our understanding of the educational outcomes of China's schooling system must remain speculative. Analysts such as Hongqi and Xiu (2012: 180) have warned against the temptation of over-stating the abilities of Chinese pupils and assuming that the country has reached an efficiency frontier. Specifically, some evidence suggests that Chinese pupils under-perform relative to their non-Chinese peers with regard to areas requiring more critical thinking or complex interpretation, for instance the interpretation of graphs. Despite an absence of nationally representative data on learning outcomes, local research involving the collection of test and background data for the purposes of production function analysis has become increasingly permitted and seems to point to an interest on the part of the Chinese government to use empirical evidence when formulating policies to strengthen educational quality (Yuhong and Zhiyong, 2008).

Some dismantling of the examinations system that schools were subject to occurred as part of the government's efforts to improve learning. In the 1990s, examinations in primary schools (Grades 1 to 6) were abolished with a view to making learning more oriented towards problem-solving and less towards rote learning (Han and Yang, 2010). Recent research suggests that some success has been achieved in getting teachers to move away from simple rote-learning and to focus more broadly on cognitive development and the creative use of information (Yee, 2010). The literature on the reforms to the vital *gaokao* examination, the examination determining access to post-school education, is less clear on what the purpose and impact of the reforms have been. There has been a significant movement away from national standardisation and towards provincial design of the *gaokao* (China has 22 provinces) (Hannum, Xuehui and Hua-Yu, 2011). Moreover, there has been a move towards examination questions requiring more critical thinking. The importance of the *gaokao* has moreover diminished somewhat as universities, in particular new private ones, have been permitted to administer their own separate entrance examinations. One risk, generally not addressed in the literature, is that these reforms would lead to inequities insofar as provinces with different standards could result in schools aspiring to different standards. Another is that universities will make inefficient admissions decisions due to the information asymmetries arising out of having different examinations in different provinces. Some of the decision-making difficulties, but not the inefficiency problem, are reduced by the fact that many universities have quotas for the number of students from different provinces. A new assessment unit established in the national ministry in 2007 has as a key objective the establishment of a national monitoring system. However, whether this means the establishment of a Prova Brasil type system is not clear. World Bank thinking would suggest that such a system is necessary to improve knowledge about which parts of the country are lagging behind educationally and whether progress is being made over time. As argued in a World Bank guide to national assessment systems (Greaney and Kellaghan, 2008), traditional examinations are unable to provide this information for a number of reasons, for instance because they are sensitive to fluctuations in who writes the examinations each year.

One policy aimed at improving labour market returns to schooling has been the consolidation of vocational training at the upper secondary level after the large-scale undermining of this training during the Cultural Revolution of the 1960s and 1970s. Vocational training constituted half of upper secondary schooling before the Cultural Revolution and after dropping to 14% of enrolments during the revolution, the ratio has been maintained at around 50%, despite large increases in the upper secondary enrolment gross enrolment ratio, for instance from 43 to 74 between 2000 and 2008 (Hannum, Xuehui and Hua-Yu, 2011: 271). Demand for vocational training at this level has been kept high by a perception amongst students and households that income returns in the labour market are high (Jing and Yu, 2006).

Better systems to promote the accountability of school principals and teachers to the authorities (but not to communities, as discussed below) has been a further area of policy focus. During the 1980s a school inspectorate was re-established after having been dismantled in the 1950s. Figures referred to by Law (2002: 584), viewed in combination with UNESCO figures, suggest that by 1998 a ratio of around 10 schools per inspector had been reached, in other words a relatively favourable ratio. The widely quoted 1993 Teachers Law[21] introduced accountability rules for teachers which appear to follow good practices in other countries, if one ignores the parts of the law dealing with teachers' duties to promote the ruling party. The performance assessment of teachers should take into account self-assessment. Teachers who perform well should be rewarded with salary promotions and greater management powers within the school. Sanctions include dismissal and could arise out of 'intentionally not accomplishing educational and teaching tasks and thus causing losses to educational and teaching work' (article 37), as well the insulting and corporal punishment of pupils.

The introduction of a legal framework for private education institutions in the 1980s was largely motivated by concerns that the public system would not be able to expand fast enough to serve the rapidly growing economy. By 2009, 10% of basic schooling enrolment was in private institutions, though most of this was in private pre-schools. It appears as if private institutions will play an increasingly important role (Law and Pan, 2009).

One type of accountability promoted strongly by many education reform proponents, including the World Bank, but not by the Chinese government, is accountability to communities. This gap in the Chinese reform programme is a serious one and is indicative of the wider problem of China's attempt to achieve economic development and modernisation without a movement towards a more democratic and participatory political system (Han and Yang, 2010). Brock (2009: 359) refers to Chinese schools operating like 'spaceships', in isolation from the surrounding community. The absence of school structures where parents can participate to some degree in the governance of the school means the absence of an accountability line that many would argue is necessary to make up for weaknesses in the line of accountability between the school and the authorities. Serious defects in this latter line of accountability in China are well documented. There is a widespread culture whereby schools fake compliance with government reform programmes because principals and teachers do not see these reforms serving their own interests. To illustrate, school principals will often maintain two school timetables, one that complies with official rules and is shown to the

[21] See http://tradeinservices.mofcom.gov.cn/en/b/1993-10-31/13698.shtml, also Law (2002).

authorities, and another one actually used in the school. The latter one will ignore curriculum innovations that the authorities consider necessary but school staff are unwilling to implement due to the work implied by the change (Walker and Qian, 2012). Assuming that these curriculum changes are changes that parents would support, the authorities could utilise pressure from parents to bring about the desired change. One might expect that China's one-child policy would make parents particularly willing to play a role in school governance, both because fewer children gives parents more time for such activities and because the policy makes it impossible to spread the risk of poor education across several children. If the only child is poorly educated, there is no recourse in the form of a well educated second or third child. Arguably, the absence of independent teacher unions represents a further inefficiency and accountability problem insofar as this leads to abuses where local government officials withhold the salaries of teachers in order to devote funds to non-personnel items that were originally not budgeted for or because they are corrupt. Such abuses appear to be widespread and can be expected to discourage teachers and make the teaching profession unattractive for youths (Law, 2002).

Turning to the matter of equity within the schooling system, we should differentiate between inequity resulting from China's strong emphasis on streaming (this inequity is easier to justify) and inequity resulting from corruption and privatisation in the system (this is less easy to justify). With regard to the first, China has for some decades promoted 'key' schools, or elite schools into which the best performing pupils would be streamed with a view to optimising the abilities of the country's future intellectual elite. In the early 1980s around 4% of secondary schools were key schools. In the mid-1980s there was an official policy shift against this type of elitism, yet since then the practice of key schools has continued in some form or another (Han and Yang, 2010).

Less defensible forms of inequity include the dropping out from compulsory level schooling of many poor children, high public funding inequities across schools, in particular depending on the province of the school, and the use of private funds by households to buy better schooling for their children within the public system. Inequality in schooling, as in society generally, is greatly exacerbated by severe restrictions on free movement within China effected through the *hukou* system, not unlike the influx control system of apartheid South Africa. This system is largely designed to restrict urbanisation by rationing access by rural workers and their families to cities.

Evidence on participation in pre-schooling in China is not easily available, but official figures suggest that China is more or less on a par with other developing countries. The gross enrolment ratio of 44 in 2008, assuming ages 4 to 6, in China is somewhat above the developing country average of 38, but below Brazil's 65 (also for ages 4 to 6) (UNESCO, 2011a). Participation in Grades 1 to 9 was made compulsory in a landmark 1986 law. By 1996, 75% of children were reaching at least the start of lower secondary schooling, or grade 7 (Jing and Yu, 2006: 259). Analysts generally believe that participation up to grade 9 is now close to 100%. Thereafter, it declines sharply, however. Hongqi and Xiu (2012), using official figures, argue that participation in upper secondary schooling, or Grades 10 to 12, is 70% in urban areas and as low as 10% in rural areas. Chengfang, Linxiu, Renfu and Scott (2009) offer an important critique of the generally accepted enrolment statistics and argue that the situation is in fact better and more equitable than what these statistics suggest. It is argued that vocational training enrolments are under-estimated. But a larger issue is the fact that despite *hukou* restrictions, migration from rural to urban areas to access better upper secondary schooling is said to be much greater than what is generally believed. Thus many rural youths do get to access upper secondary schooling by migrating to cities. Chengfang *et al* (2009) estimate that in fact around 25% of rural youths access this level of schooling and that official participation rates for urban youths must consequently be adjusted downward to some degree.

The 1986 compulsory schooling law banned school fees up to the end of grade 9. Yet contraventions of this ban have been widespread and loopholes have been exploited to a large degree. To illustrate, many public schools demand that parents contribute fees to cover bicycle parking at the school (Law, 2002). Brock (2009: 454) argues that there is a widespread belief amongst government officials that fees are a necessity, even for the poor, as they signal the commitment of households to schooling and hence make it easier for the education authorities to argue for better public funding. Chengfang *et al* (2009) argue that at the upper secondary level fees reach levels that seriously undermine educational equity. These fees often amount to over 500 USD per year. If one adds to this the fact that the opportunity cost is high as employment opportunities for youths leaving lower secondary schooling are relatively good, it becomes clear why the government should consider low participation in upper secondary schooling a serious ‘bottleneck’ in the education system (Jing and Yu, 2006). Jing and Yu (2006) see this problem as a manifestation of poorly designed decentralisation, where powers are devolved to a lower level but in the absence of sufficiently clear policy parameters that can protect the national interest. All schooling up to grade 12 was devolved from the national to the provincial level in the mid-1980s, but whilst it was made clear that schooling up to grade 9 would be compulsory, there was little guidance relating to upper secondary schooling.

Policy creates the impression that the central government will subsidise provincial authorities sufficiently for national school participation goals, where they exist, to be pursued. However, in reality existing funding systems are clearly inadequate. The demand for funding, in particular as far as poorer households is concerned, is aggravated by the fact that upper secondary schooling often occurs in boarding schools. In the more rural western provinces, even lower secondary schools tend to have boarding facilities. Grants to support attendance of the poor in upper secondary schools exists, but only around 8% of eligible pupils receive these grants and, what is even worse, they reportedly only cover around 2% of the costs incurred by households (Chengfang *et al*, 2009). One area where there has been considerable success in targeting educational resources towards the poor is textbooks. Since 2000, free textbooks have been widely distributed to all pupils in the west and to rural pupils in the east of the country (Brock, 2009). That education should be seriously under-funded in China seems to be confirmed by aggregate spending figures. During the 1990s the ratio of public education spending over GDP, which was already low at the start of the period, declined according to Law (2002). Government has acknowledged this as a problem and committed itself to raising the ratio, for instance to 4.0% by 2000, a fairly average level by developing country standards. However, by 2000 a level of just 2.5% had been reached (Brock, 2009: 455). There is clearly a public accountability problem, evidenced by the fact that the UNESCO international database contained no ratios of education spending to GDP, or some other related statistic, for any year after 1999, in 2012 when the database was queried[22]. What makes this apparent under-investment in human capital more remarkable is the fact that China's investment in physical capital in recent years, as measured by gross capital formation as a proportion of income, is above the world average and at least twice as high as what has been seen in the three other countries dealt with above (World Bank, 2011b: 398). For Brock (2009), a key factor underlying China's under-investment in human capital is the highly corrupt and nepotistic practices characterising the distribution of funds intended for schools. There is a political unwillingness to direct more funds into a corrupt system, but at the same time there is an unwillingness or an inability to root out the corruption causing the problem in the first place. Notwithstanding these problems, overall the funding of schools has been improving in recent years, according to Ding (2012), and analysis suggests that lack of funds is no longer the largest reason why poor pupils drop out of school. Instead, more educational problems such as the poor quality of schooling and the difficulty of the curriculum play a larger role.

It appears as if one factor that kept the demand for private schools relatively low is that the illicit purchasing of advantages within the public system by the rich is widespread. In one city studied by Law (2002), 20% of pupils enrolled in academically elite schools were there not because they had complied with the strict academic entrance requirements, but rather because parents had paid an additional fee, often reaching 8,000 USD, to secure a place in the school for their child. Law (2002: 591), in lamenting the inability of the authorities to enforce educational rules, argues that in China 'there is an emerging belief in the magic power of law to rectify irregularities in education and society.'

As is the case in Brazil and South Africa, generous subsidies exist for students wishing to become teachers (Hannum *et al*, 2011). The length of training for different teachers varies greatly, with upper secondary teachers undergoing four years of post-school training, whilst primary school teachers only need to complete grade 12 at a vocational secondary school specialising in the training of teachers[23].

At the post-school level, China has seen a phenomenal increase in enrolments since the early 1990s, from around 4 million to 31 million in 2010[24]. The gross enrolment ratio increased from 3 to 26 during the same period. The number of graduates in the labour market has been further boosted by the arrival every year of thousands of Chinese graduates returning home after completing their tertiary education outside the country. The OECD estimated the annual figure to be 20,000 (Gallagher, Hasan, Canning and Newby, 2009: 6). China now has the second-highest number of researchers of any country, after the United States. According to Hannum *et al* (2011), China's extraordinary expansion of its tertiary education sector was partly made possible as a result of marketisation of this sector. In around 1990 two key policy changes occurred. Firstly, tertiary institutions were granted permission to supplement their grants received from the state with fees charged to students. In effect, 3% to 8% of overall costs could be covered by fees from students. Secondly, it became possible for institutions to admit certain students who did not fulfil the academic entrance

[22] Online data-querying facility of UNESCO: UIS, consulted 2012.
[23] See http://www.edu.cn/20010101/21925.shtml.
[24] Online data-querying facility of UNESCO: UIS, consulted 2012. Also Shujie, Bin, Fang and Jianling (2010: 841) and Hannum *et al* (2011: 269).

requirements, if they paid privately for the full cost of their education. The establishment of a few completely private universities contributed to the trend, but only to a limited degree. One further critical explanation for the very rapid increase in enrolments, an increase that reached 30% per year for the three years 1998 to 2000, was that the previous student to lecturer ratio had been remarkably low. There was thus spare capacity, at least in terms of lecturers, to accommodate a rapid increase. The student to lecturer ratio went up from a low level of 3:1 to 17:1 following the 1990 changes (Jing and Yu, 2006: 257). Access to tertiary education became considerably less equitable in the process, despite the introduction of targeted loans and scholarships for exceptionally performing students from poor households (Shujie *et al*, 2010; Gallagher *et al*, 2009: 6). For instance, there is evidence that amongst students in elite key universities, the proportion of students from poorer backgrounds has declined (Hannum *et al*, 2011: 276). Hannum *et al* (2011: 276) argue that reforms to bring about more pupil-centred learning have had the unintended consequence of making academic achievement amongst the socio-economically disadvantaged relatively more difficult. They also argue that second-chance schemes favour the rich.

Whether the rapid expansion in post-school education has led to greater inequalities in the quality of education is a matter that is not resolved. What has occurred is that earnings differentials amongst graduates have increased, but it is not clear to what extent this reflects an increasingly competitive labour market, where earnings have become more sensitive to the quality of human capital, and to what extent it reflects greater inequality with respect to the quality of education (Zhong, 2011). The OECD's review of China's tertiary education sector emphasises the need for greater national government involvement in benchmarking and enforcing standards across China's universities to counteract increasing incoherence within the system, in particular at the undergraduate level. Moreover, given China's rising importance as a world power, benchmarking against major existing quality frameworks elsewhere is regarded as a necessary improvement. Quality benchmarks, according to the OECD, are likely to incentivise more innovative teaching practices and greater awareness of employer needs. In one key respect China's universities are rather unusual: in the 1980s permanent employment tenure, also known as the 'iron rice bowl', for university lecturers was stopped, resulting in a situation where all academic staff were placed on temporary contracts. How this arrangement impacts on the quality of university teaching is not clear (Gallagher *et al*, 2009: 6-10, 37).

15 CONCLUSIONS FOR PART III

The shift from focussing just on grade attainment to both grade attainment and the quality of educational outcomes may not be complete, as suggested by the 2010 UNDP Human Development Report discussed in section 13, yet education policymakers appear to have made a critical shift in this direction. Of the four countries examined in section 14, Chile has the longest history of monitoring learning outcomes according to standardised indicators, but Brazil has accumulated several years of experience in this area since national standardised tests were first introduced in the mid-1990s. South Africa has recently embarked on the use of standardised testing for accountability purposes. China, however, has followed a different and arguably less scientific and systematic approach to monitoring learning outcomes. An established tradition of examinations at various levels of the schooling system appears to have maintained a strong focus on what pupils learn. However, this has not permitted a proper understanding of how the quality of basic education in China compares to that of other countries and, perhaps more importantly, whether this is improving over time within China. Moreover, the partial dismantling of some of the national standardisation of the examinations in recent years, added to the fact that at least in 2012 there appeared to be no plans for a Prova Brasil-type national assessment system, raises the possibility that monitoring and thus accountability will become more elusive. In particular, the absence of a sufficiently standardised national system of quality monitoring could result in greater human capital inequality across the country. This poses a serious challenge, especially given concerns around rising inequality in China with respect to income.

The scepticism towards standardised assessments expressed by Education International, the world federation of teacher unions, seems not to be influencing current policy strongly, at least not in the four developing countries discussed above (three of which have teacher unions). Why would teacher unions, which are said to be so influential, not have undermined the development of national assessment systems to a greater degree? Judging from the available literature, a likely explanation seems to be that despite the positions held by Education International, unions at the country level are not strongly opposed to standardised assessment, at least not in Chile, Brazil and South Africa. It is possible that they see the need for them and realise that opposing the assessment systems and the associated accountability mechanisms would evoke public criticism of the unions.

What does the discussion in the preceding sections suggest should receive special attention within the economics of education field? Innovative but very different approaches taken in both Brazil and Chile to use information for institutional accountability purposes in the post-school sector point to a potentially important means of addressing quality problems in this sector. The economic literature has paid considerable attention to the impact of different accountability interventions at the school level, but such literature dealing with post-school education is scarce. Here it seems as if the policymakers have been a step ahead of the economists. Economic analysis of the impact of standardised information emerging from university testing systems (Brazil) and publicly available information on labour market returns linked to specific post-school institutions (Chile), in both cases generated as part of government programmes, seems warranted. The fact that the Brazilian programme appears not to be seen as a violation of academic freedoms is significant for a country such as South

Africa, where sensitivities around academic freedom are strong though institution-driven quality control mechanisms appear not to be having the desired impact.

Corruption within the education sector is a further area that may warrant a stronger focus. The World Bank has paid considerable attention to corruption relating to financial flows, specifically through its public expenditure tracking survey (PETS) methodology[25]. However, education corruption is a broad topic that includes nepotism in the appointment of managers, a problem acknowledged explicitly in South Africa's policies, and the non-payment of teachers by corrupt administrators, a problem found in China with its absence of teacher unions and mechanisms for parent representation. A significant move is Transparency International's new focus on reporting on corruption within the education sector. The organisation's 2011 report on corruption in the primary school sector in South Africa is an indication of this move (Døssing, Mokeki and Weideman, 2011). Heyneman (2004b) offers a rare and apparently useful framework for understanding corruption in the broad sense within the education sector. At the highest level, three areas in which corruption occurs can be identified: the selection of students for promotion to a higher level; the accreditation of education institutions, giving them the right to operate; and the supply of public funds and goods within the education system. The World Bank's PETS work has focused on the last of these three. One advantage with using corruption as an organising framework for one's analysis of education sector inefficiencies is that corruption is an emotive topic that appears to be receiving increasing public attention. It is thus a framework that can assist in attracting a wider audience for one's analysis.

Targets for expanding the education sector, specifically post-school education enrolments in Brazil and South Africa, were found to be unrealistic and poorly informed by trends in other countries. This could simply be a manifestation of ritualistic planning, where targets are intentionally aspirational rather than realistic and empirically informed. However, even if intended to be ritualistic, unrealistic targets carry risks as they could impact in perverse ways on, for instance, institutional planning. There is little guidance in the economic literature for policymakers who do wish to specify realistic sector expansion targets. As explained with reference to Glewwe's (2002) discussion of rates of return and education planning in Part I of the dissertation, part of the challenge lies in a more careful application of existing economic evidence to the question of education sector growth. But part of the challenge is also to make use of analysis such as that presented in Part II to establish benchmarks for the spread of resources and efforts across the different levels of the education system.

What is it in the existing economic research on education systems that should be underlined when researchers communicate with policymakers? As highlighted by Bruns, Filmer and Patrinos (2011: 62), the evidence that the multitude of in-service teacher training programmes (and other types of quality enhancement interventions) have had a positive impact on learning is extremely weak. Government in-service teacher training programmes tend not to include any impact evaluation component, perhaps because it is incorrectly assumed that the designed activities must inevitably lead to improvements. Even Brazil's ambitious programme appears to be severely limited with respect to an impact evaluation element. This is a serious gap given the obvious importance of teacher capacity in the school improvement process. Where

[25] See for instance Reinikka and Smith (2004).

impact evaluations have been conducted in developing countries, they have tended to emerge out of experimental interventions driven by non-government stakeholders. One message that economists need to emphasise is that impact evaluations are important, with regard to in-service training but also other policy areas, partly because the little evidence that exists indicates that many interventions that seem to do the right thing at face value, in fact result in very little change. What this points to is a need to challenge more explicitly the received wisdom in education policymaking circles around what interventions work. Perhaps because it is so difficult to get even the basics, such as large-scale in-service training programmes, in place when it comes to improving a schooling system, the perception has arisen that achieving this is enough. A step forward would therefore be to promote the understanding that though having the basics in place is indeed important, this is not enough. Those basics need to be introduced with much caution and with an awareness that appearances may deceive.

There is perhaps something to be said for a better packaging of the economics of education messages within a clearer political economy framework. Given that some of the resistance to accountability reforms, in particular on the part of teacher unions, uses a political economy argument, it may be of value for economists to deal specifically with this aspect of the policy debate. This is what the World Bank analyst Benveniste (2002: 106) does when he argues in favour of standardised testing systems because such systems encourage not just institutional accountability, but also government accountability to the electorate. In this sense, he argues, these testing systems can be considered a necessary element of the welfare state.

Part III can be said to make a methodological contribution towards comparative education approaches, a research area that Psacharopoulos (1990) has argued needs fixing. What Part III has done is to attempt to evaluate policies through the comparison of several education systems within an economic framework dealing with external efficiency, internal efficiency and equity. These kinds of analyses are rare. In particular, there seems to be very little work on evaluating China's education system holistically, at least in the literature available in English, despite this education system being the world's second-largest in terms of the annual intake of pupils at the primary school level[26].

[26] Despite China's population still exceeding that of India's in 2008, India's primary school intake figure of 32 million was almost twice the 17 million of China (UNESCO, 2011a). In 2008, the total primary and secondary enrolments for the two countries were virtually equal.

16 OVERALL CONCLUSIONS

Each of the three parts of the dissertation has its own conclusion. Part I moreover had conclusions for four sub-sections dealing (1) the theory of economic growth, (2) the measurement of human capital and national development, (3) existing empirical models of growth and (4) ways of communicating the economic knowledge to education policymakers. In Part II, the conclusion assessed what had been learnt about the prioritisation of education levels from a series of analyses of UNESCO Institute for Statistics (UIS) country-level data. The conclusion to Part III discussed the views of education policymakers and their receptiveness to research produced in the field of economics. These conclusions were aimed partly at answering a number of questions that had been posed. At the start of Part I, an over-arching question for the dissertation as a whole was proposed: *What is the utility of the economic growth literature and models using country-level data for education policymakers and to what extent have these policymakers made use of the available evidence?*

The aim in this overall conclusion is not simply to repeat or summarise what has already been said in the separate conclusions that have appeared above. Rather, the aim is to draw from those conclusions to answer the over-arching question reproduced above.

So what is the utility of the economic growth literature for education policymakers? This literature undoubtedly serves as a vehicle for emphasising that how countries organise their education systems matters hugely for future income and well-being. But is this not something everyone knows, one may ask. Is education's vital role in human development not self-evident? As discussed in the introduction to section 2, many, though not all, philosophy of education lecturers who educate future schoolteachers are not comfortable with the notion that schooling is an activity that serves future welfare improvements. And the philosophers of education are not alone. The discomfort is not so surprising if one considers that it was only in the 1960s that economists themselves started to pay serious attention to the causal link between human capital growth and welfare improvements. Using the evidence we now have on this link to underscore the importance of education for economic and social development seems like something we cannot do too much of.

It can be argued that the economic growth literature has since around 2005 become even more interesting and important for education policymakers, due to the shift in emphasis from years of schooling attained as a measure of human capital, to average test score results. The earlier growth models using years of schooling did point to the vital role of education. Mankiw, Romer and Weil (1991), who can be said to have pioneered the use of a years of schooling variable in a growth regression concluded that without this variable, the growth model would not make sense, it would display too many internal inconsistencies. The work of Sala-i-Martin, Doppelhofer and Miller (2004) aimed at ranking the importance of 67 variables that seemed important explanatory variables for growth, concluded that initial level of enrolment at the primary school level was the second-most important variable.

However, the finding of Hanushek and Woessman (for instance 2009) that the introduction of an education quality measure in a growth model renders any education quantity measure essentially insignificant, has justifiably excited a range of organisations and individuals concerned with how to improve education systems. One can speculate that much of the excitement arises out of the realisation that such basic things as the ability to read and a person's numeracy explain to a large degree such a complex phenomenon as economic growth. If our only measure of human capital were years of schooling, there would be far less certainty around the components of this input. Years of schooling could represent many things: basic skills, but also socialisation, or a person's range of subject knowledge. Hanushek and Woessman's work has narrowed down the essence of human capital. It is about basic skills. The implication is that the central challenge for education systems is to maximise these skills. Socialisation, or the creation of social capital, and curriculum diversity, are likely to also be important components of human capital, but it is implied that these things flow almost naturally from the initial investment in the achievement of basic skills.

But some of the excitement must arise not just from the apparent simplicity of what lies at the core of human capital, but also the ease with which this can be measured, and the fact that we now have mature international testing programmes that have been improved over many years of trial and error. Getting education systems to rally around improvements in internationally standardised measures of literacy and numeracy is a relatively uncomplicated idea, even if realising the improvements remains a highly complex matter. The value of a degree of simplicity on the outcomes end of the education production process for policymakers, and voters and parents, should not be under-estimated. Evidence of the kind put forward by Hanushek and Woessman helps to counteract the notion that education outcomes are too elusive to allow for measurement and simple goals, a notion that is widely embraced, for instance by the world federation for teacher unions, Education International. Of course education outcomes are not a simple matter and one can expect an increasingly nuanced, and less excited, response to the Hanushek and Woessman findings as these findings become entrenched as mainstream thinking within policy circles. Hanushek and Woessman (2009: 22) do themselves provide an initial exploration of an important policy question, namely the degree to which equity in learning outcomes matters, relative to average performance. Is it a matter of 'rocket scientists or education for all', they asked. This complexity, and others, are likely to receive more attention in future. Moreover, as argued in Part I of the dissertation, it must be acknowledged that though Hanushek and Woessman's (2009) work represents a major step forward, relative to previous work where the distinction between participation in schooling and learning outcomes was blurred, Hanushek and Woessman (2009) remains not fully conclusive due to unavoidable data problems. Specifically, it will be necessary to repeat their work as time passes and we obtain longer historical series of statistics on country-level educational performance.

The more innovative parts of the dissertation dealt with how theory and empirical findings, expressed within economics texts, can be communicated to policymakers and a wider audience beyond researchers. The assumption here was that the utility of research for policymakers needs to be assessed not just in terms of the original research, but also in terms of the means by which researchers and others communicate research findings to governments. Existing modes of communication were critically assessed and new devices were developed that could serve to underscore and, to a reasonable degree, simplify the messages of the economists.

To a large degree, the work by Hanushek and Woessman served as a point of departure, given that they have been exceptionally successful at bridging the often serious gap between researchers and education policymakers. Two new devices, building on Hanushek and Woessman, stand out. One is the expansion of Hanushek and Woessman's list of 77 countries with comparable measures of educational quality, to a list of 113 countries, through the inclusion of more developing country data drawn from an African and a Latin American testing programme which had not been considered by Hanushek and Woessman (section 3.3 in Part I). The expanded list adds value partly because it allows a larger number of national governments to see where they are located in the global rankings of educational outcomes, and what opportunities exist for improving income per capita through better schooling.

The second device put forward in the dissertation (section 5.5 in Part I) is an Excel-based 'growth simulator' tool aimed at informing policymakers not only of the likely impact of better educational quality on income (something that Hanushek and Woessman had dealt with), but also how important improving education is relative to other interventions, such as reducing ethnolinguistic fractionalisation, promoting trade openness and combating malaria. The dynamic nature of the Excel tool allows users to compare countries rather easily, something that would be more difficult in a more static report. The tool confirmed the relative importance of investing in educational improvement. But the relative strengths of other interventions are noteworthy too, even for education policymakers, given the potential impact of education policies on non-education variables. For instance, in South Africa improving educational quality emerged as the second-most promising stimulator of growth, after improving life expectancy. This points to the importance of health education in South Africa's schools, including effective education around the prevention of HIV/AIDS. In South Africa, reducing linguistic fractionalisation also emerged as important. Language is a matter that seems to receive inadequate attention in South Africa's policies. In examining these dynamics, use was made of what appears to be the only dataset available in South Africa on people's (self-reported) competencies in the country's eleven official languages.

One area that is of particular importance for education policymakers, and which one might expect to be dealt with in the growth literature, is the balancing of budgets and policy attention across the levels of the education system, in particular the primary, secondary and tertiary levels. Despite the availability of a couple of interesting studies focussing on the relationship between levels of educational attainment in the population and income, this area has received little attention. What seemed missing was a systematic analysis of patterns in the historical enrolment and spending figures in education systems across the world, by education level, and correlations between these patterns and income. For instance, have faster growing economies prioritised resources differently across the levels of the education system?

The country-level data of UNESCO's Institute for Statistics (UIS) seemed to offer the opportunity to produce a novel analysis in this regard. Part II of the dissertation was largely devoted to presenting such an analysis. One key finding was that faster growing economies do indeed display different education investment patterns. Their gross enrolment ratios across all education levels are higher, as are their per student spending levels. Interestingly, given scepticism around the benefits of tertiary education spending in rich countries amongst economists, the analysis in Part II found that faster growing rich economies have spent more on each student. It was also found that one cannot conclude unequivocally from the data that Sub-Saharan Africa has over-invested in primary schooling, relative to secondary and tertiary education. This is important given the strong presence in the literature of the argument that aid donors have forced Africa to neglect the secondary and tertiary levels.

In Part II it was also demonstrated how comparing one country's prioritisation across the education levels to global patterns can help to underscore imbalances in the education system. The demonstration made use of the South African situation and confirmed what has been argued elsewhere in different ways, namely that South Africa has over-prioritised secondary schooling and under-prioritised tertiary education. In line with the dissertation's concern around ways of communicating patterns in the data to policymakers, Part II paid considerable attention to devising new graphs able to represent prioritisation across education levels.

To what extent have education policymakers made use of the evidence emerging from the field of economics? Given that economists claim to offer evidence-based insights not only into the size and nature of the causal relationship between educational improvement and average income, but also on ways in which the inner workings of the education system can be improved, one might expect education policymakers to pay attention. The question was addressed in Part III by looking at the receptiveness of education policymakers to the economics of education literature in general (not just the literature on growth models). There is a sense amongst the economists that policymakers should be more receptive. The examination in Part III of actual policies in developing countries, and in Part I (section 5.2) of the literature dealing with the 'research-policy nexus', or communication between researchers and policymakers, suggests that the complaint of the economists is partly justified, but partly not. On the one hand, there are powerful ideological tendencies in the education sector, as well as less noble tendencies such as rent-seeking, which would resist the penetration of economic thinking. On the other, researchers could pay more attention to communicating single research findings in easily readable formats, and to formulating meta-analyses of the vast body of literature, partly in order to distinguish the more credible research from weak research.

Part III drew a distinction between education researchers with close commitments to the teaching profession, and perhaps teacher unions, and education policymakers. The former tend to be far less accepting of the economics literature than the latter. The way the economic findings are communicated should depend partly on what audience one is aiming at. The examination of four developing country education systems in Part III suggests that the economic literature could be better communicated to policymakers, but not that there is any noteworthy resistance to this literature amongst the policymakers. Both education economists and policymakers seem to share the opinion that basic learning outcomes, or cognitive skills, need to be measured through properly standardised testing systems, both for monitoring and accountability purposes. Where economists could be more insistent, is in explaining the implications of the emergence of more and better impact evaluations focussing on interventions to improve schooling. These evaluations have brought to the fore an inconvenient truth, namely that interventions such as in-service teacher training programmes that policymakers place so much faith in tend not to make any significant positive difference to the learning outcomes of pupils (World Bank, 2003: 114, 123). A better awareness of this is likely to bring about more caution in the design of these interventions, and a greater use of formal impact evaluation techniques.

Lastly, education economists, for understandable reasons, tend to want to plan their research work around the microdata they have access to. There is nothing wrong with this. But economists could strengthen their own impact on education policymaking if they paid more attention to glaring anomalies in government plans, and how these anomalies could be corrected through the application of economic principles and the available evidence. One area apparently plagued by these anomalies is the expansion of the post-school sector. This was seen in both South Africa and Brazil. What is needed here, in particular in the case of South Africa, is a clearer acknowledgement of the technologies by which the capital intensity (in particular human capital intensity) of tertiary education could be reduced, the role of institutional accountability and how the public and private sectors can complement each other.

BIBLIOGRAPHY

Aghion, P. & Howitt, P. (1998). *Endogenous growth theory*. Cambridge (Massachusetts): MIT Press.

Aghion, P. & Howitt, P. (2009). *The economics of growth*. Cambridge (Massachusetts): MIT Press.

Agrawala, S. (1998). Context and early origins of the Intergovernmental Panel on Climate Change. *Climatic Change*, 39: 605-620.

Alesina, A., Devleeschauwer, A., Easterly, W. & Kurlat, S. (2003). Fractionalization. *Journal of Economic Growth*, 8: 155-194.

Allgood, S., Bosshardt, W., Van der Klaauw, W. & Watts, M. (2012). Is economics coursework, or majoring in economics, associated with different civic behaviours? *Research in economic education*, 43(3): 248-268.

Arrow, K. (1962). The economic implications of learning by doing. *Review of Economic Studies*, 29: 155-173.

Bailey, T. (2010). *The research-policy nexus: Mapping the terrain of the literature*. Wynberg: Centre for Higher Education Transformation.

Banerjee, A., Galiani, S., Levinsohn, J., McLaren, Z. & Woolard, I. (2007). *Why has unemployment risen in the new South Africa*. Cambridge: NBER. Available from: <http://www.nber.org/papers/w13167.pdf?new_window=1> [Accessed October 2007].

Bangay, C. & Blum, N. (2010). Education responses to climate change and quality: Two parts of the same agenda? *International Journal of Educational Development*, 30: 359-368.

Barat, L.M. (2006). Four malaria success stories: How malaria burden was successfully reduced in Brazil, Eritrea, India and Vietnam. *American Journal of Tropical Medicine and Hygiene*, 74(1): 12-16.

Barro, R.J. & Lee, J.-W. (2000). International data on educational attainment (dataset). Cambridge: Harvard University. Available from: <http://www.cid.harvard.edu/ciddata/ciddata.html>

Barro, R.J. & Lee, J.-W. (2001). International data on educational attainment: Updates and implications. *Oxford Economic Papers*, 3: 541-563.

Barro, R.J. & Sala-i-Martin, X. (2003). *Economic growth*. Cambridge: MIT Press.

Baum, C.F. (2006). *An introduction to modern econometrics using Stata*. College Station: Stata.

Becker, G.S. (1993). Nobel lecture: The economic way of looking at behavior. *The Journal of Political Economy*, 101(3): 385-409.

Benavot, A. & Riddle, P. (1988). The expansion of primary education 1870-1940. *Sociology of Education*, 61(3): 191-210.

Benhabib, J. & Spiegel, M.M. (1994). The role of human capital in economic development: Evidence from aggregate cross-country data. *Journal of Monetary Economics*, 34: 143-173.

Bennell, P. (2006). Using and abusing rates of return: A critique of the World Bank's 1995 education sector review. *International Journal of Educational Development*, 16(3): 235-248.

Benveniste, L. (2002). The political structuration of assessment: Negotiating state power and legitimacy. *Comparative Education Review*, 46(1): 89-118.

Bertrand, O. (2004). *Planning human resources: Methods, experiences and practices.* Paris: IIEP. Available from: <http://www.unesco.org> [Accessed December 2006].

Bils, M. & Klenow, P.J. (2000). Does schooling cause growth? *Journal of Economic Growth*, 90(5): 1160-1183.

Binder, T.R. & Contreras, D. (2008). *Tournaments, gift exchanges, and the effect of monetary incentive for teachers: The case of Chile*. Santiago: University of Chile. Available from: <http://www.depeco.econo.unlp.edu.ar/cedlas/ien/pdfs/meeting2008/papers/rau-contreras.pdf> [Accessed February 2012].

Blalock, H.M. (1979). *Social statistics*. Singapore: McGraw-Hill.

Bof, A.M. (2004). Distance learning and teacher training in Brazil. *International Review of Research in Open and Distance Learning*, 5(1).
Boissiere, M. (2004). *Rationale for public investments in primary education in developing countries.* Washington: World Bank. Available from: <http://www.worldbank.org> [Accessed November 2006].
Bolt, K., Matete, M. & Clemens, M. (2002). *Manual for calculating adjusted net saving*. Washington: World Bank. Available from: <http://www.worldbank.org/> [Accessed April 2011].
Branson, N. & Leibbrandt, M. (2013). Education quality and labour market outcomes in South Africa. Paris: OECD. Available from: <http://www.oecd-ilibrary.org/> [Accessed October 2013].
Brazil: Ministério da Educação (2001). *Plano Nacional de Educação*. Brasilia. Available from: <https://www.planalto.gov.br/ccivil_03/leis/leis_2001/l10172.htm> [Accessed March 2012].
Brazil: Ministério da Educação (2010). *Plano Nacional de Educação (projeto de lei)*. Brasilia. Available from: <http://portal.mec.gov.br> [Accessed March 2012].
Breton, T.R. (2011). The quality vs. the quantity of schooling: What drives economic growth? *Economics of Education Review*, 30(4): 765-773.
Brock, A. (2009). Moving mountains stone by stone: Reforming rural education in China. *International Journal of Educational Development*, 29: 454-462.
Browne, L. & Alstrup, P. (2006). *What exactly is the Labour Force Survey?* London: National Statistics. Available from: <http://www.docstoc.com/docs/30807058/What-exactly-is-the-Labour-Force-Survey> [Accessed November 2012].
Bruns, B. (2010). *Achieving world class education in Brazil: The next agenda*. Washington: World Bank. Available from: <http://www.iepecdg.com.br/uploads/livros/1012achieving_world.pdf> [Accessed October 2011].
Bruns, B., Filmer, D. & Patrinos, H.A. (2011). *Making schools work: New evidence on accountability reforms*. Washington: World Bank. Available from: <http://siteresources.worldbank.org/EDUCATION/Resources/278200-1298568319076/makingschoolswork.pdf> [Accessed October 2011].
Burger, R. & Du Plessis, S. (2006). *Examining the robustness of competing explanations of slow growth in African countries*. Nottingham: University of Nottingham. Available from: <http://driver-support.eu/economics/credit/research/papers/cp0602.pdf> [Accessed October 2010].
Burger, R.P. & Teal, F.J. (2013). *Measuring the option value of education*. Stellenbosch: University of Stellenbosch. Available from: <http://ideas.repec.org/e/pbu98.html> [Accessed October 2013].
Carnoy, M. & Rhoten, D. (2002). What does globalization mean for educational change? A comparative approach. *Comparative Education Review*, 46(1): 1-9.
Case, A. & Deaton, A. (1999). School inputs and educational outcomes in South Africa. *Quarterly Journal of Economics*, 114(3): 1047-1084.
Chengfang, L., Linxiu, Z., Renfu, L. & Scott, R. (2009). Development challenges, tuition barriers, and high school education in China. *Asia Pacific Journal of Education*, 29(4): 503-520.
Chile: Congreso Nacional (2011). *Ley calidad y equidad*. Santiago. Available from: <http://www.leychile.cl> [Accessed April 2012].
Cobb, C.W. & Douglas, P.H. (1928). A theory of production. *The American Economic Review*, 18 (Supplement): 139-165.
Cohen, D. & Soto, M. (2007a). Growth and human capital: Good data, good results. *Journal of Economic Growth*, 12(1): 51-76.
Cohen, D. & Soto, M. (2007b). Education database (dataset). Barcelona: Instituto de Análisis Económico. Available from: <http://soto.iae-csic.org/Data.htm> [Accessed January 2012].

CONFEMEN (2012). *Synthèse régionale des résultats PASEC - Document de travail.* Dakar. Available from: <http://www.confemen.org/wp-content/uploads/2012/03/Synthese_PASEC_VII-VIII-IX_final.pdf> [Accessed December 2012].

Coulombe, S., Tremblay, J.-F. & Marchand, S. (2004). *International Adult Literacy Survey: Literacy scores, human capital and growth across fourteen OECD countries*. Ottawa: Statistics Canada. Available from: <http://www.nald.ca/library/research/oecd/oecd.pdf> [Accessed October 2012].

Crouch, L (1999). Province and race historical enrolment. Pretoria. [Excel file named *province and race historical enrolment.xls*].

Crouch, L. (2005). Political economy, incentives and teachers' unions. In Emiliana Vegas (ed.), *Incentives to improve teaching: Lessons from Latin America*. Washington: World Bank: 103-150.

Crouch, L. & Vinjevold, P. (2006). *South Africa: Access before quality, and what to do now?* Granada: Universidad de Granada. Available from: <http://www.ugr.es/~recfpro/rev101ART6ing.pdf> [Accessed June 2010].

Curle, A. (1969). *The professional identity of the educational planner*. Paris: UNESCO: IIEP. Available from: <http://unesdoc.unesco.org/images/0007/000762/076210eo.pdf> [Accessed December 2008].

D'Alessandro, S., Luzzati, T. & Morroni, M. (2008). *Transition paths towards economic and environmental sustainability: Feasible scenarios and policy implications*. Pisa: Università di Pisa. Available from: <http://www.esee2009.si/papers/D%20Alessandro-Transition_paths.pdf> [Accessed May 2011].

Davies, R. & Thurlow, J. (2009). *Formal-informal economy linkages and unemployment in South Africa*. Cape Town: Economic Research Southern Africa. Available from: <http://www.econrsa.org/wkshops/cge/session2-formal-informal-economy-linkages-unemployment-south-africa.pdf> [Accessed December 2010].

DeCanio, S.J. (2003). *Economic models of climate change: A critique*. New York: Palgrave Macmillan.

Delandshere, G. & Petrosky, A. (2004). Political rationales and ideological stances of the standards-based reform of teacher education in the US. *Teaching and Teacher Education*, 20: 1-15.

Denison, E.F. (1974). *Accounting for United States economic growth 1929-1969.* Washington: The Brookings Institution.

De Villiers, P. (2009). *The cost of fiscal subsidies to higher education students in South Africa: A comparison between 2000 and 2006*. Stellenbosch: University of Stellenbosch. Available from: <http://ideas.repec.org/p/sza/wpaper/wpapers85.html> [Accessed September 2012].

Di Tella, R., MacCulloch, R.J. & Oswald, A.J. (2003). The macroeconomics of happiness. *The Review of Economics and Statistics*, 85(4): 809-827.

Ding, Y. (2012). The problems with access to compulsory education in China and the effects of the policy of direct subsidies to students. *Chinese Education and Society*, 45(1): 13-21.

Dixit, A.K. & Stiglitz, J.E. (1977). Monopolistic competition and optimum product diversity. *The American Economic Review*, 67(2): 297-308.

Døssing, H., Mokeki, L. & Weideman, M. (2011). *Mapping transparency, accountability and integrity in primary education in South Africa*. Berlin: Transparency International. Available from: <http://www.transparency.org> [Accessed October].

Downs, A. (1957). An economic theory of political action in a democracy. *Journal of Political Economy*, 65(2): 135-150.

Du Rand, G., Van Broekhuizen, H. & Von Fintel, D. (2011). *Numeric competence, confidence and school quality in the South African wage function: Towards understanding pre-labour market discrimination.* Stellenbosch: University of Stellenbosch. Available from: <http://ideas.repec.org/p/sza/wpaper/wpapers140.html> [Accessed February 2012].

Duflo, E., Glennester, R. & Kremer, M. (2006). *Using randomization in development economics research: A toolkit*. Washington: NBER. Available from: <http://www.nber.org/papers/t0333> [Accessed March 2013].
Easterly, W. & Levine, R. (1997). Africa's growth tragedy: Policies and ethnic divisions. *Quarterly Journal of Economics*, 112(4): 1023-1250.
Easterly, W. & Levine, R. (2001). What have we learned from a decade on empirical research on growth? It's not factor accumulation: Stylized facts and growth models. *The World Bank Economic Review*, 15(2): 177-219.
Education International (2011). *Policy paper on education: Building the future through quality education*. Brussels. Available from: <http://www.ei-ie.org> [Accessed October 2011].
Emmerij, L. (2002). Aid as a flight forward. *Development and change*, 33(2): 247-260.
Ervik, A.O. (2003). IQ and the wealth of nations (book review). *The Economic Journal*, 113: 406-408.
Fearon, J.D. (2003). Ethnic and cultural diversity by country. *Journal of Economic Growth*, 8: 195-222.
Fedderke, J.W. & Luiz, J.M. (2008). Does human capital generate social and institutional capital? Exploring evidence from South African time series data. *Oxford Economic Papers*, 60: 649-682.
Fernandes, R. (2007). *Índice de Desenvolvimento da Educação Básica (Ideb)*. Brasília: Ministério da Educação. Available from: <http://www.publicacoes.inep.gov.br/detalhes.asp?pub=4121#> [Accessed November 2009].
Finkel, S.E. & Ernst, H.R. (2005). Civic education in post-apartheid South Africa: Alternative paths to the development of political knowledge and democratic values. *Political Psychology*, 26(3): 333-364.
Flannery, T. (2005). *The weather makers: The history and future impact of climate change*. London: Penguin.
Frankel, M. (1962). The production function in allocation and growth: A synthesis. *The American Economic Review*, 52(5): 995-1022.
Friedman, M. (1955). The role of government in education. In R. A. Solo (ed.), *Economics and the public interest*. New Brunswick: Rutgers University Press.
Gallup, J.L., Mellinger, A.D. & Sachs, J.D. (2001). General measures of geography (dataset). Cambridge: Center for International Development at Harvard University. Available from: <http://www.cid.harvard.edu/ciddata/geographydata.htm> [Accessed July 2010].
Galston, W.A. (2001). Political knowledge, political engagement and civic education. *Annual Review of Political Science*, 4(217-234).
Gallagher, M., Hasan, A., Canning, M. & Newby, H. (2009). *OECD reviews of tertiary education: China*. Paris: OECD. Available from: <http://www.oecd.org/dataoecd/42/23/42286617.pdf> [Accessed May 2011].
Gilead, T. (2012). Education and the logic of economic progress. *Journal of the Philosophy of Education*, 46(1): 113-131.
Ginsburg, M.B. & Gorostiaga, J.M. (2001). Relationships between theorists/researchers and policy makers/practitioners: Rethinking the two-cultures thesis and the possibility of dialogue. *Comparative Education Review*, 45(2): 173-196.
Glewwe, P. (1996). The relevance of standard estimates of rates of return to schooling for education policy: A critical assessment. *Journal of Development Economics*, 51(2): 267-290.
Greaney, V. & Kellaghan, T. (2008). *Assessing national achievement levels in education*. Washington: World Bank. Available from: <http://siteresources.worldbank.org/EDUCATION/Resources/278200-1099079877269/547664-1099079993288/assessing_national_achievement_level_Edu.pdf> [Accessed March 2010].

Griliches, Z. (2002). Productivity: Measurement problems. In J. Eatwell (ed.), *The new Palgrave: A dictionary of economics*. New York: Houndmills.
Gujarati, D.N. (2003). *Basic econometrics*. New York: McGraw-Hill.
Gustafsson, M. (2011). *The when and how of leaving school: The policy implications of new evidence on secondary schooling in South Africa*. Stellenbosch: University of Stellenbosch. Available from: <http://ideas.repec.org/p/sza/wpaper/wpapers137.html> [Accessed March 2011].
Gustafsson, M. (2012a). *More countries, similar results: A nonlinear programming approach to normalising test scores needed for growth regressions*. Stellenbosch: Stellenbosch University. Available from: <http://ideas.repec.org/p/sza/wpaper/wpapers164.html> [Accessed August 2012].
Gustafsson, M. (2012b). *The gap between school enrolments and population in South Africa: Analysis of the possible explanations*. Stellenbosch: University of Stellenbosch. Available from: <http://ideas.repec.org/p/sza/wpaper/wpapers176.html> [Accessed December 2012].
Gustafsson, M. (2012c). *Incentives for teachers within the salary system.* [Unpublished report].
Gustafsson, M. & Mabogoane, T. (2012). South Africa's economics of education: A stocktaking and an agenda for the way forward. *Development Southern Africa*, 29(3): 351-364.
Gustafsson, M. & Patel, F. (2009). *Managing the teacher pay system: What the local and international data are telling us*. Stellenbosch: Stellenbosch University. Available from: <http://ideas.repec.org/p/sza/wpaper/wpapers99.html> [Accessed June 2010].
Gustafsson, M., Van der Berg, S., Shepherd, D. & Burger, C. (2010). *The costs of illiteracy in South Africa*. Stellenbosch: University of Stellenbosch. Available from: <http://ideas.repec.org/s/sza/wpaper.html> [Accessed July 2010].
Han, M. & Yang, X. (2010). Educational assessment in China: Lessons from history and future prospects. *Assessment in Education: Principles, Policy & Practice*, 8(1): 5-10.
Hannum, E., Xuehui, A. & Hua-Yu, S.C. (2011). Examinations and educational opportunity in China: Mobility and bottlenecks for the rural poor. *Oxford Review of Education*, 37(2): 267-305.
Hanushek, E.A. & Woessman, L. (2007). *The role of school improvement in economic development.* Washington: National Bureau of Economic Research. Available from: <http://papers.nber.org/papers/w12832.pdf?new_window=1> [Accessed June 2007].
Hanushek, E.A. & Woessman, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, 46(3): 607-668.
Hanushek, E.A. & Woessman, L. (2009). *Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation*. Washington: National Bureau of Economic Research. Available from: <http://www.nber.org/papers/w14633.pdf?new_window=1> [Accessed March 2009].
Hanushek, E.A. & Wößmann, L. (2007). *Education quality and economic growth*. Washington: World Bank. Available from: <http://siteresources.worldbank.org/EDUCATION/Resources/278200-1099079877269/547664-1099079934475/Edu_Quality_Economic_Growth.pdf> [Accessed October 2009].
Harley, K. & Wedekind, V. (2004). Political change, curriculum change and social formation, 1990 to 2002. In Linda Chisholm (ed.), *Changing class: Education and social change in post-apartheid South Africa*. Pretoria: HSRC. Available from: <http://www.hsrcpublishers.ac.za> [Accessed December 2006].
Hawaii: Department of Education (2003). *Reading and mathematics AYP starting points, intermediate goals, annual measurable objectives.* Honolulu. Available from: <http://arch.k12.hi.us/PDFs/nclb/AYP_Reading-Math.pdf > [Accessed March 2013].
Heertje, A. (2008). Schumpeter, Joseph Alois (1883-1950). In Steven N. Durlauf and Lawrence E. Blume (ed.), *The new Palgrave: A dictionary of economics*. New York: Houndmills.

Heston, A., Summers, R. & Aten, B. (2006). Penn World Table version 6.2 (dataset). Philadelphia: Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania. Available from: <http://pwt.econ.upenn.edu/php_site/pwt_index.php> [Accessed June 2007].

Heston, A., Summers, R. & Aten, B. (2011). Penn World Table version 7.0 (dataset). Philadelphia: Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania. Available from: <http://pwt.econ.upenn.edu/php_site/pwt_index.php> [Accessed June 2007].

Heyneman, S.P. (2004a). International education quality. *Economics of Education Review*, 23: 441-452.

Heyneman, S.P. (2004b). Education and corruption. *International Journal of Educational Development*, 24(6): 637–648.

Hodgson, D. & Watkins, S.C. (1997). Feminists and neo-Malthusians: past and present alliances. *Population and Development Review*, 23(3): 469.

Hongqi, C. & Xiu, C.C. (2012). Principal professional development in China: Challenges, opportunities, and strategies. *Peabody Journal of Education*, 87(2): 178-199.

Hong-Sang, J. & Thorbecke, E. (2001). *The impact of public education expenditure on human capital, growth and poverty in Tanzania and Zambia: A general equilibrium approach.* Washington: International Monetary Fund. Available from: <http://ideas.repec.org/p/imf/imfwpa/01-106.html> [Accessed July 2009].

Hu, W. (2012). Thirty years of education in China: Between change and no change. *Chinese Education and Society*, 45(1): 84-94.

Hyman, D.N. (1999). *Public finance: A contemporary application of theory to policy*. Fort Worth: The Dryden Press.

IIEP: SACMEQ (2004). SACMEQ II (2000 dataset). Harare.

IEA (2004). TIMSS 2003 (dataset). Chestnut Hill: Boston College. Available from: <http://timss.bc.edu/timss2003.html> [Accessed July 2007].

Inbar, D.E. (1996). *Planning for innovation in education*. Paris: IIEP. Available from: <http://www.unesco.org> [Accessed December 2006].

Inglehart, R. & Welzel, C. (2005). *Nation-level mean scores on traditional/secular-rational and survival/self-expression values dimensions*. Madrid: World Values Survey. Available from: <http://www.worldvaluessurvey.org/> [Accessed March 2011].

Ironmonger, D. (2001). *Household production and the household economy*. Melbourne: University of Melbourne. Available from: <http://www.economics.unimelb.edu.au/downloads/wpapers-00-01/759.pdf> [Accessed December 2010].

Jackson, T. (2009a). *Prosperity without growth: Economics for a finite planet.* London: Earthscan.

Jackson, T. (2009b). *Prosperity without growth? The transition to a sustainable economy*. London: Sustainable Development Commission. Available from: <http://www.sd-commission.org.uk/data/files/publications/prosperity_without_growth_report.pdf> [Accessed May 2011].

Jamison, E.A., Jamison, D.T. & Hanushek, E.A. (2007). The effects of education quality on income growth and mortality decline. *Economics of Education Review*, 26(6): 771-788.

Jing, L. & Yu, Z. (2006). Educational expansion and shortages in secondary schools in China: The bottle neck syndrome. *Journal of Contemporary China*, 15(47): 255-274.

Jones, G. & Schneider, W.J. (2006). Intelligence, human capital and economic growth: A Bayesian averaging of classical estimates (BACE) approach. *Journal of Economic Growth*, 11: 71-93.

Kastellec, J.P. & Leoni, E.L. (2007). Using graphs instead of tables in political science. *Perspectives on politics*, 5(4): 755-771.

Kennedy, P. (2008). *A guide to econometrics*. Malden: Blackwell.

KPMG (2011). *China's 12th five-year plan: An overview.* Beijing: KPMG. Available from: < http://www.kpmg.com/CN/en/IssuesAndInsights/ArticlesPublications/Publicationseries/5-

years-plan/Documents/China-12th-Five-Year-Plan-Overview-201104.pdf> [Accessed March 2013].
Kraak, A. & Press, K., eds. (2008). *Human resources development review 2008*. Available from: <http://www.hsrcpress.ac.za> [Accessed March 2008].
Krueger, A.B. & Lindahl, M.B. (2001). Education for growth: Why and for whom? *Journal of Economic Literature*, 39(4): 1101-1136.
Krueger, D. & Kumar, K.B. (2004). Skill-specific rather than general education: A reason for Europe-US growth differences? *Journal of Economic Growth*, 9(2): 167-207.
Kuznets, S. (1980). Driving forces of economic growth: What can we learn from history? *Review of World Economics*, 116(3): 409-431.
Kyriacou, G.A. (1991). *Level and growth effects of human capital: A cross-country study.* New York: New York University. Available from: <http://ideas.repec.org/p/cvs/starer/91-26.html> [Accessed September 2012].
Law, W.-W. (2002). Legislation, education reform and social transformation: The People's Republic of China experience. *International Journal of Educational Development*, 22: 579-602.
Law, W.-W. & Pan, S.-Y. (2009). Game theory and educational policy: Private education legislation in China. *International Journal of Educational Development*, 29(2): 227-240.
Layard, R., Clark, A. & Senik, C. (2012). The causes of happiness and misery. In John Helliwell, Richard Layard and Jeffrey Sachs (eds.), *World happiness report*. New York: The Earth Institute. Available from: <http://earth.columbia.edu/articles/view/2960> [Accessed October 2013].
Lin, J. & Chang, H.-J. (2012). Should industrial policy in developing countries conform to comparative advantage or defy it? A debate between Justin Lin and Ha-Joon Chang. *Development Policy Review*, 27(5): 483-502.
Lipsey, R.G. & Lancaster, K. (1956). The general theory of second best. *The Review of Economic Studies*, 24(1).
Lewin, K.M. (2011). *Taking targets to task revisited: How indicators of progress on access to education can mislead*. Falmer: University of Sussex. Available from: <http://www.create-rpc.org/pdf_documents/PTA54.pdf> [Accessed July 2011].
Louw, M., Van der Berg, S. & Yu, D. (2006). *Educational attainment and intergenerational social mobility in South Africa*. Stellenbosch: University of Stellenbosch. Available from: <http://ideas.repec.org/p/sza/wpaper/wpapers23.html> [Accessed June 2010].
Loveless, T. (2005). Test-based accountability: The promise and the perils. *Brookings Papers on Education Policy*, 8: 7-45.
Lucas, R.E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22: 3-42.
Lucas, R.E. (2003). Macroeconomic priorities. *American Economic Review*, 93(1): 1-14.
Lutz, W., Goujon, A., Samir, K.C. & Sanderson, W. (2007). Reconstructions of population by age, sex and level of educational attainment for 120 countries for 1970-2000. *Vienna Yearbook of Population Research*. 193-235.
Lutz, W., Cuaresma, J.C. & Sanderson, W. (2008). The demography of educational attainment and economic growth. *Science*, 319 (1047-1048).
Makuwa, D.K. (2010). Mixed results in achievement. *IIEP Newsletter*, XXVIII(3). Available from: <http://www.iiep.unesco.org> [Accessed October 2010].
Mandel, E. (2002). Marx. In J. Eatwell (ed.), *The new Palgrave: A dictionary of economics.* New York: Houndmills.
Mankiw, N.G., Romer, D. & Weil, D.N. (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics*, 107(2): 407-443.
Manuelli, R.E. & Seshadri, A. (2010). *Human capital and the wealth of nations*. Madison: University of Wisconsin-Madison. Available from: <http://www.econ.wisc.edu/~aseshadr/working_pdf/humancapital.pdf> [Accessed September 2012].

Masanjala, W.H. & Papageorgiou, C. (2008). Rough and lonely road to prosperity: A reexamination of the sources of growth in Africa using Bayesian model averaging. *Journal of Applied Econometrics*, 23: 671-682.
Matola, M. (2012). *Highs and lows of our competitiveness.* Johannesburg: The Sunday Independent <http://www.iol.co.za/sundayindependent/highs-and-lows-of-our-competitiveness-1.1384893#.UTuLKdZmAhU> [Accessed March 2013].
McCloskey, D.N. (1994). *Knowledge and persuasion in economics.* Cambridge: Cambridge University Press.
McCloskey, D. (2008). How to buy, sell, make, manage, produce, transact, consume with words. In Edeward M. Clift (ed.), *How language is used to do business: Essays on the rhetoric of economics*. Lewiston: Mellen Press.
McEwan, P.J., Urquiola, M. & Vegas, E. (2008). School choice, stratification and information on school performance: Lessons from Chile. *Economia*, 8(2): 1-42.
McMahon, W. (1982). *Spillover effects and option values in education*. Urbana-Champaign: University of Illinois. Available from: <https://www.ideals.illinois.edu/bitstream/handle/2142/27172/spillovereffects881mcma.pdf?sequence=1> [Accessed October 2013].
Meckes, L. & Carrasco, R. (2010). Two decades of SIMCE: an overview of the national assessment system in Chile. *Assessment in education: Principles, policy and practice*, 17(2): 233-248.
Mingat, A. & Tan, J.-P. (1998). *The mechanics of progress in education: Evidence from cross-country data*. Washington: World Bank. Available from: <http://www.worldbank.org> [Accessed November 2003].
Mingat, A. & Tan, L.-P. (2003). On the mechanics of progress in primary education. *Economics of Education Review*, 22: 455–467.
Mourshed, M., Chijioke, C. & Barber, M. (2010). *How the world's most improved school systems keep getting better.* New York: McKinsey & Company. Available from: <http://ssomckinsey.darbyfilms.com/reports/schools/How-the-Worlds-Most-Improved-School-Systems-Keep-Getting-Better_Download-version_Final.pdf> [Accessed February 2010].
Mueller, D.C. & Stratmann, T. (2003). The economic effectsof democratic participation. *Journal of Public Economics*, 87(9-10): 2129-2155.
Mullis, I.V.S., Martin, M.O., Gonzalez, E.J. & Chrostowski, S.J. (2004). *TIMSS 2003 international mathematics report*. Chestnut Hill: Boston College. Available from: <http://timss.bc.edu> [Accessed January 2008].
Mullis, I.V.S., Martin, M.O., Gonzalez, E.J. & Kennedy, A.M. (2003). *PIRLS 2001 international report.* Chestnut Hill: Boston College. Available from: <http://timss.bc.edu> [Accessed April 2007].
Mullis, I.V.S., Martin, M.O. & Foy, P. (2008). *TIMSS 2007 international mathematics report*. Chestnut Hill: Boston College. Available from: <http://timss.bc.edu> [Accessed May 2009].
Mullis, I.V.S., Martin, M.O., Foy, P. & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill: Boston College. Available from: <http://timss.bc.edu> [Accessed December 2012].
Mullis, I.V.S., Martin, M.O., Kennedy, A.M. & Foy, P. (2007). *PIRLS 2006 international report.* Chestnut Hill: Boston College. Available from: <http://timss.bc.edu> [Accessed April 2007].
Nelson, R.R. & Phelps, E.S. (1966). Investment in humans, technological diffusion, and economic growth. *The American Economic Review*, 56(2): 69-75.
OCDE (2009). *La educación superior en Chile*. Paris. Available from: <http://www.opech.cl/educsuperior/politica_educacion/la_es_en_chile_ocde.pdf> [Accessed April 2012].
OECD (2000). *Literacy in the information age: The International Adult Literacy Survey.* Paris. Available from:

<http://www.oecd.org/edu/highereducationandadultlearning/41529765.pdf> [Accessed December 2012].
OECD (2001). PISA 2000 database (dataset). Paris. Available from: <http://www.oecd.org/document/48/0,3343,en_32252351_32236159_33668528_1_1_1_1,00.html> [Accessed July 2007].
OECD (2004a). *Learning for tomorrow's world: First results from PISA 2003.* Paris. Available from: <http://www.oecd.org/document/55/0,2340,en_32252351_32236173_33917303_1_1_1_1,00.html> [Accessed April 2007].
OECD (2004b). PISA 2003 database (dataset). Paris. Available from: <http://www.oecd.org/document/0/0,3343,en_32252351_32236173_33694144_1_1_1_1,00.html> [Accessed July 2007].
OECD (2005). *Learning a living: First results of the Adult Literacy and Life Skills Survey*. Paris. Available from: <http://www.oecd.org/education/educationeconomyandsociety/34867438.pdf> [Accessed December 2012].
OECD (2007). *PISA 2006: Science competencies for tomorrow's world*. Paris Available from: <http://www.oecd.org/dataoecd/30/17/39703267.pdf> [Accessed May 2008].
OECD (2009). *PISA 2006 technical report*. Paris Available from: <http://www.oecd.org/dataoecd/0/47/42025182.pdf> [Accessed May 2010].
OECD (2010a). *The high cost of low educational performance: The long-run economic impact of improving PISA outcomes*. Paris Available from: <http://www.oecd.org/dataoecd/11/28/44417824.pdf> [Accessed May 2010].
OECD (2010b). *PISA 2009 results: What students know and can do*. Paris.
OECD (2012). *Strong performers and successful performers in education: Lessons from PISA for the United States*. Paris Available from: <http://www.pearsonfoundation.org/oecd/downloads/Brazil_strong_performers_US.pdf> [Accessed July 2012].
O'Neill, B.C., MacKellar, F.L. & Lutz, W. (2005). *Population and climate change*. Cambridge: Cambridge University Press.
Parandekar, S., Amorim, É. & Welsh, A. (2008). *Prova Brasil: Building a framework to promote measureable progress in learning outcomes*. Washington: World Bank. Available from: <http://siteresources.worldbank.org/INTENBREVE/Newsletters/21809358/Mar08_121_Prova_EN_FINAL.pdf> [Accessed November 2009].
Patrinos, H.A. & Psacharopoulos, G. (2011). *Education: Past, present and future global challenges*. Washington: World Bank. Available from: <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1799163> [Accessed May 2011].
Pearce, D. (2002). An intellectual history of environmental economics. *Annual Review of Energy and the Environment*, 27: 57-81.
Pritchett, L. & Filmer, D. (1997). *What education production functions really show: A positive theory of education expenditures*. Washington: World Bank. Available from: <http://www.worldbank.org> [Accessed March 2007].
Pritchett, L. (2001). Where has all the education gone? *The World Bank Economic Review*, 15(3): 367-391.
Pritchett, L., Samji, S. & Hammer, J. (2013). *It's all about MeE: Using structured experiential learning ("e") to crawl the design space.* Washington: Center for Global Development. Available from: <http://papers.ssrn.com/> [Accessed February 2014].
Pritchett, L., Woolcock, M. & Andrews, M. (2012). Looking like a state: Techniques of persistent failure in state capability for implementation. *Journal of Development Studies*, 49(1): 1-18.
Psacharopoulos, G. (1989). Why education reforms fail: A comparative analysis. *International Education Review*, 35(2): 179-195.
Psacharopoulos, G. (1990). Comparative Education: From Theory to Practice, or Are You A:\neo.* or B:*. ist? *Comparative Education Review*, 34(3): 369-380.

Psacharopoulos, G. (1994). Returns to investment in education: A global update. World *Development*, 22(9): 1324-1343.
Psacharopoulos, G. (1996). Economics of education: A research agenda. Economics *of Education Review*, 15(4): 339-344.
Psacharopoulos, G., Tan, J.-P. & Jimenez, E. (1986). *Financing education in developing countries: An exploration of policy options*. Washington: World Bank. Available from: <http://www.eric.ed.gov/PDFS/ED281800.pdf> [Accessed January 2012].
Psacharopoulos, G. & Arriagada, A.-M. (1986). *The educational attainment of the labor force: An international comparison*. Washington: World Bank. Available from: <http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2005/09/01/000112742_20050901145133/Rendered/PDF/edt38.pdf> [Accessed October 2012].
Psacharapoulos, G. & Patrinos, H.A. (2002). *Returns to investment in education: A further update*. Washington: World Bank. Available from: <http://www.worldbank.org> [Accessed November 2006].
Reinikka, R. & Smith, N. (2004). *Public expenditure tracking surveys in education*. Paris: IIEP. Available from: <http://prof.estat.com/cgi-bin/ft/233033148987?class=IIEP_PDF_pubs&page=F166&estat_url=http://www.unesco.org/iiep/PDF/pubs/Reinikka.pdf> [Accessed November 2006].
Resnik, J. (2006). International organisations, the "education-economic growth" black box, and the development of world education culture. *Comparative Education Review*, 50(2): 173-195.
Reuveny, R. & Li, Q. (2003). Economic Openness, Democracy, and Income Inequality: An Empirical Analysis. *Comparative Political Studies*, 36(5): 575-601.
Rezende, M. (2010). The effects of accountability on higher education. *Economics of Education Review*, 29(5): 842-586.
Roeder, P.G. (2001). Ethnolinguistic Fractionalization (ELF) Indices, 1961 and 1985 (dataset). San Diego: University of California. Available from: <http//:weber.ucsd.edu\~proeder\elf.htm> [Accessed February 2013].
Romer, P.M. (1986). Increasing returns and long-run growth. *Journal of Political Economy*, 94(5): 1002-1037.
Romer, P.M. (1989). *Human capital and growth: Theory and evidence.* Washington: National Bureau of Economic Research. Available from: <http://www.nber.org/papers/w3173> [Accessed April 2007].
Rosling, H., Rosling Rönnlund, A. & Rosling, O. (2005). New software brings statistics beyond the eye. In Enrico Giovannini (ed.), *Statistics, knowledge and policy: Key indicators to inform decision making*. Paris: OECD: 522-530.
Sacks, D.W., Stevenson, B. & Wolfers, J. (2010). *Subjective well-being, income, economic development and growth.* Washington: NBER. Available from: <http://www.nber.org/papers/w16441.pdf?new_window=1> [Accessed January 2011].
Sala-i-Martin, X., Doppelhofer, G. & Miller, R.I. (2004). Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach. *The American Economic Review*, 94(4): 813-835. Data and programming code used for this paper available from <http://www.nhh.no/Default.aspx?ID=3075> [Accessed August 2011].
Saliola, F. & Seker, M. (2011). *Total factor productivity across the developing world*. Washington: World Bank. Available from: <http://www.enterprisesurveys.org/~/media/FPDKM/EnterpriseSurveys/Documents/Enterprise%20Notes/Productivity-23.pdf> [Accessed September 2012].
Samir, K.C., Barakat, B., Goujon, A., Skirbekk, V., Sanderson, W. & Lutz, W. (2010). Projections of population by level of educational attainment, age, and sex for 120 countries 2005-2050. *Demographic Research*, 22: 383-472.
Samoff, J. (1999). Education sector analysis in Africa: Limited national control and even less national ownership. *International Journal of Educational Development*, 19(4-5): 249-272.

Schlotter, M., Schwerdt, G. & Woessman, L. (2010). *Econometric methods for causal evaluation of education policies and practices: A non-technical guide*. Bonn: IZA. Available from: <http://ftp.iza.org/dp4725.pdf> [Accessed April 2013].

Schneider, D. (2011). *The labor share: A review of theory and evidence*. Berlin: Humboldt-Universität. Available from: <http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2011-069.pdf> [Accessed September 2012].

Schofer, E. & Meyer, J.W. (2005). The worldwide expansion of higher education in the twentieth century. *American Sociological Review,* 70(6): 898-920.

Schultz, T.W. (1961). Investment in human capital. *The American Economic Review*, 51(1): 1-17.

Sen, A. (1999). *Development as freedom*. New York: Alfred A. Knopf.

Shujie, Y., Bin, W., Fang, S. & Jianling, W. (2010). The impact of higher education expansion on social justice in China: A spatial and inter-temporal analysis. *Journal of Contemporary China*, 19(67): 837-854.

Smith, A. (1991). *The wealth of nations.* London: Everyman's Library.

Smith, A. (2010). *Lectures on jurisprudence*. Glasgow: University of Glasgow. Available from: <http://files.libertyfund.org > [Accessed April 2011].

Solow, R.M. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics*, 70(1): 65-94.

Solow, R.M. (2001). What have we learned from a decade on empirical research on growth? Applying growth theories across countries. *The World Bank Economic Review*, 15(2): 283-288.

South Africa: Council on Higher Education (2010). *Report on the national review of academic and professional programmes in education*. Pretoria. Available from: <http://www.che.ac.za/> [Accessed August 2011].

South Africa: Department of Basic Education (2010a). *Delivery agreement for outcome 1: Improved quality of basic education.* Pretoria. Available from: <http://www.education.gov.za/DeliveryAgreement/tabid/446/Default.aspx> [Accessed March 2012].

South Africa: Department of Basic Education (2010b). *Call for comments on action plan to 2014: Towards the realisation of Schooling 2025 (Government Notice 752)*. Pretoria. Available from: <http://www.gov.za> [Accessed January 2012].

South Africa: Department of Basic Education (2011a). *Integrated strategic planning framework for teacher education and development in South Africa*. Pretoria. Available from: <http://www.education.gov.za> [Accessed June 2011].

South Africa: Department of Basic Education (2011b). *Action Plan to 2014: Towards the realisation of Schooling 2025*. Pretoria. Available from: <http://www.education.gov.za> [Accessed June 2012].

South Africa: Department of Education (1995). *White paper on education and training*. Pretoria. Available from: <http://www.info.gov.za/whitepapers/1995/education1.htm> [Accessed March 2012].

South Africa: Department of Education (2002). *Revised national curriculum statement Grades R to 9 (schools): Life orientation.* Pretoria. Available from: <http://www.education.gov.za> [Accessed June 2011].

South Africa: Department of Education (2006). *2005 education investment review.* Pretoria.

South Africa: Department of Higher Education and Training (2011). *Policy on the minimum requirements for teacher education qualifications (Government Notice 583)*. Pretoria. Available from: <http://www.gov.za> [Accessed April 2012].

South Africa: Department of Higher Education and Training (2012). *Green paper for post-school education and training*. Pretoria. Available from: <http://www.dhet.gov.za> [Accessed March 2012].

South Africa: Department of Higher Education and Training (2013). *Statistics on post-school education and training in South Africa: 2011*. Pretoria. Available from: <http://www.dhet.gov.za> [Accessed October 2013].

South Africa: Financial and Fiscal Commission (2012). *Submission for the 2013/14 division of revenue.* Halfway House. Available from: <http://www.ffc.co.za> [Accessed September 2013].

South Africa: HSRC (2005). HSRC Client Survey October 2004 (dataset). Pretoria.

South Africa: National Student Financial Aid Scheme (2012). *Annual report*. Cape Town. Available from: <http://www.nsfas.org.za> [Accessed September 2013].

South Africa: National Planning Commission (2011). *National development plan: Vision for 2030*. Pretoria. Available from: <http://npconline.co.za> [Accessed March 2012].

South Africa: National Planning Commission (2012). *National development plan 2030: Our future - make it work.* Pretoria. Available from: <http://npconline.co.za> [Accessed December 2012].

South Africa: Presidency (2009). *Together doing more and better: Medium term strategic framework.* Pretoria.

South African Reserve Bank (2009). *Quarterly bulletin: March 2009*. Pretoria. Available from: <http://www.reservebank.co.za> [Accessed April 2009].

Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics*, 87(3): 355-374.

StataCorp (2011). *Stata multivariate statistics reference manual release 12*. College Station: Stata Press.

Statistics South Africa (2001). *A survey of time use*. Pretoria. Available from: <http://www.statssa.gov.za> [Accessed December 2010].

Statistics South Africa (2010). *Gross domestic product: Third quarter: 2010 (Statistical release P0441)*. Pretoria. Available from: <http://www.statssa.gov.za> [Accessed December 2010]. [Reports in the same series from other years also consulted.]

Statistics South Africa (2011). *Mid-year population estimates, 2011*. Pretoria. Available from: <http://www.statssa.gov.za> [Accessed October 2005] [Reports in the same series from other years also consulted.].

Statistics South Africa (2012). General Household Survey 2011 (dataset). Pretoria. [Datasets in the same series from earlier years also consulted.]

Stern, N. (2006a). What is the economics of climate change? *World Economics*, 7(2): 1-10.

Stern, N. (2006b). *Stern review on the economics of climate change*. London: United Kingdom: HM Treasury. Available from: <http://www.hm-treasury.gov.uk/independent_reviews/stern_review_economics_climate_change/stern_review_report.cfm> [Accessed May 2007].

Stiglitz, J.E. (1975). The theory of "screening", education and the distribution of income. *The American Economic Review*, 65(3): 283-300.

Stiglitz, J.E., Sen, A. & Fitoussi, J.-P. (2009). *Report by the Comission on the Measurement of Economic Performance and Social Progress*. Paris. Available from: <http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf> [Accessed September 2009].

Sydsæter, K. & Hammond, P. (2002). *Essential mathematics for economic analysis*. Harlow: Pearson Education.

Swan, T.W. (1956). Economic growth and capital accumulation. *The Economic Record*, 32(334-361).

Taylor, S. & Spaull, N. (2013). *The effects of rapidly expanding primary school access on effective learning: The case of Southern and Eastern Africa since 2000*. Stellenbosch: University of Stellenbosch. Available from: <http://ideas.repec.org/p/sza/wpaper/wpapers177.html> [Accessed March 2013].

Taut, S., Cortés, F., Sebastian, C. & Preiss, D. (2009). Evaluating school and parent reports of the national student achievement testing system (SIMCE) in Chile: Access, comprehension and use. *Evaluation and Program Planning*, 32(2): 129-137.

Taut, S., Santelices, M.V. & Stecher, B. (2011). *Validation of the Chilean national teacher evaluation system*. Santiago: Sociedad Chilena de Políticas Públicas. Available from: <http://www.sociedadpoliticaspublicas.cl/archivos/BLOQUEM/Docentes_y_Directivos_Escolares/Teacher_Evaluation_System.pdf> [Accessed June 2012].

Teal, F. (2011). Higher education and economic development in Africa: A review of channels and interactions. *Journal of African Economies*, 20(Supplement 3): iii50-iii79.
Thorn, W. (2009). *International adult literacy and basic skills surveys in the OECD region*. Paris: OECD. Available from: <http://www.oecd-ilibrary.org/docserver/download/5ksb2zx31xmx.pdf?expires=1354455651&id=id&accname=guest&checksum=E49F1116125330C996D9F5BBA7B18A1C> [Accessed December 2012].
Torney-Purta, J. (2002). The school's role in developing civic engagement. *Applied Developmental Science*, 6(4): 203-212.
Transparency International (2011). *Corruption perceptions index 2011*. Berlin. Available from: <http://www.transparency.org> [Accessed March 2012].
UNDP (2010). *Human Development Report 2010: The real wealth of nations: Pathways to human development*. New York. Available from: <http://www.undp.org> [Accessed April 2011].
UNESCO (1997). *International Standard Classification of Education*. Paris. Available from: <http://www.unesco.org> [Accessed January 2005].
UNESCO (2003). *Education indicators: Technical guidelines*. Paris. Available from: <http://www.uis.unesco.org> [Accessed January 2005].
UNESCO (2005). *Education for All global monitoring report 2005: The quality imperative.* Paris Available from: <http://www.unesco.org/education/gmr_download/chapter2.pdf> [Accessed May 2008].
UNESCO (2008). *Primer reporte: Los aprendizajes de los estudiantes de América Latina y el Caribe.* Santiago: OREALC/UNESCO. Available from: <http://unesdoc.unesco.org/images/0016/001606/160660s.pdf> [Accessed March 2010].
UNESCO (2011a). *Education for All global monitoring report 2011: The hidden crisis: Armed conflict and education*. Paris Available from: <http://www.unesco.org> [Accessed August 2011].
UNESCO (2011b). *International standard classification of education 2011*. Paris. Available from: <http://www.uis.unesco.org/Education/Documents/UNESCO_GC_36C-19_ISCED_EN.pdf> [Accessed November 2012].
United Kingdom: UK Commission for Employment and Skills (2011). *Working futures 2010-2020*. Coventry. Available from: <http://www2.warwick.ac.uk/fac/soc/ier/research/wf/evidence-report-41-working-futures-2010-2020.pdf> [Accessed July 2013].
United Nations (1987). *Report of the World Commission on Environment and Development. New York*. Available from: <http://worldinbalance.net/pdf/1987-brundtland.pdf> [Accessed April 2011].
United Nations (1995). *1993 Demographic Yearbook*. New York Available from: <http://unstats.un.org/unsd/demographic/products/dyb/default.htm> [Accessed November 2012].
United Nations Statistics Division (2009a). *System of national accounts 2008*. New York Available from: <http://unstats.un.org/unsd/nationalaccount/SNA2008.pdf> [Accessed December 2010].
United Nations Statistics Division (2009b). *Detailed structure and explanatory notes: ISIC Rev.3.1*. New York Available from: <http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=17&Lg=1> [Accessed May 2009].
United States: Social Security Administration (1996). *The Boskin Commission Report*. Washington. Available from: <http://www.ssa.gov/history/reports/boskinrpt.html> [Accessed April 2011].
Van der Berg, S. (2007). Apartheid's enduring legacy: Inequality in education. *Journal of African Economies*, 16(5): 849-880.
Vandenbussche, J., Aghion, P. & Meghir, C. (2006). *Growth, distance to frontier and composition of human capital.* Journal of Economic Growth, 11: 97-127.
Veloso, F., Pessoa, S., Henriques, R. & Giambiagi, F., eds. (2009). *Educação Básica no Brasil*. Rio de Janeiro: Elsevier.

Verspoor, A. & Bregman, J. (2007). *At the crossroads: Challenges for secondary education in Africa*. Washington: World Bank. Available from: <http://www.worldbank.org> [Accessed April 2008].
Victor, P.A. & Rosenbluth, G.A. (2007). Managing without growth. *Ecological Economics*, 61: 492-504.
Villegas-Reimers, E. (2003). *Teacher professional development: An international review of the literature*. Paris: IIEP. Available from: <http://unesdoc.unesco.org/images/0013/001330/133010e.pdf> [Accessed March 2012].
Walker, A. & Qian, H. (2012). Reform disconnection in China. *Peabody Journal of Education*, 87(2): 162-177.
Walker, M. (2011). *PISA 2009 plus results*. Camberwell: Australian Council for Educational Research. Available from: <https://mypisa.acer.edu.au/images/mypisadoc/acer_pisa%202009%2B%20international.pdf> [Accessed December 2011].
Weiss, C.H. (1979). The many meanings of research utilization. *Public Administration Review*, 39(5): 426-431.
West, E. (1996). *Education vouchers in practice and principle: A world survey*. Washington: World Bank. Available from: <http://siteresources.worldbank.org/EDUCATION/Resources/278200-1099079877269/547664-1099079934475/547667-1135281552767/Education_Vouchers_WorldSurvey.pdf> [Accessed July 2008].
World Bank (1978). *World Development Report 1978*. Washington Available from: <http://www.worldbank.org> [Accessed September 2012].
World Bank (2008). *Global purchasing power parities and real expenditures: 2005 International Comparison Program*. Washington Available from: <http://siteresources.worldbank.org/ICPINT/Resources/icp-final.pdf> [Accessed December 2010].
World Bank (2010). World governance indicators (dataset). Washington. Available from: <http://info.worldbank.org/governance/wgi/pdf/wgidataset.xls> [Accessed June 2010].
World Bank (2011a). *Learning for all: Investing in people's knowledge and skills to promote development*. Washington. Available from: <http://www-wds.worldbank.org> [Accessed January 2012].
World Bank (2003). *World Development Report 2004: Making services work for poor people*. Washington Available from: <http://www.worldbank.org> [Accessed March 2012].
World Bank (2011b). *World Development Report 2012: Gender equality and development*. Washington Available from: <http://www.worldbank.org> [Accessed March 2012].
World Economic Forum (2012). *The Global Competitiveness Report 2012-2013*. Geneva Available from: <http://www.weforum.org> [Accessed November 2012].
World Health Organization (2011). Life tables for WHO member states (dataset). Geneva. Available from: <http://www.who.int/healthinfo/statistics/mortality_life_tables/en/> [Accessed April 2011].
World Values Survey Association (2009). World Values Survey 1981-2008 official aggregate v.20090901 (dataset). Madrid Available from: <www.worldvaluessurvey.org> [Accessed February 2009].
WWF (2010). *Living Planet Report 2010*. Gland: WWF. Available from: <www.panda.org> [Accessed March 2011].
Yee, H.P.J. (2010). Understanding the ecology of education reforms: Comparing the perceptions of parents and secondary teachers and students in China. *Peabody Journal of Education*, 87(2): 267-282.
Yuhong, D. & Zhiyong, Z. (2008). Impact assessment of education investment in China: Evidence from the basic education in Western Areas Project. *Chinese Education and Society*, 41(5): 3-7.
Zhong, H. (2011). Returns to higher education in China: What is the role of college quality? *China Economic Review*, 22: 260-275.

ONLINE DATA-QUERYING FACILITIES CONSULTED

UNESCO: UIS. Education statistics. Montreal. Available from: <http://www.uis.unesco.org> [Last accessed February 2012]

United Nations Statistics Division. National accounts estimates of main aggregates. New York. Available from: <http://data.un.org> [Last accessed February 2012]

United Nations Statistics Division. World population prospects: The 2008 revision. New York. Available from: <http://data.un.org> [Last accessed February 2012]

World Bank. World Bank DataBank. Available from: <http://databank.worldbank.org> [Last accessed February 2012]