

https://dspace.org/

Evaluating development effectiveness assessing and comparing the impact od education intervention in South Africa

Besharati, Neissan Alessandro

2016

http://hdl.handle.net/10539/23721

Besharati, Neissan Alessandro (2016) Evaluating development effectiveness assessing and comparing the impact of education intervention in South Africa, University of the Witwatersrand, Johannesburg, <http://hdl.handle.net/10539/23721>

Downloaded from a DSpace repository. https://dspace.org/

WITS SCHOOL OF GOVERNANCE
UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Neissan Alessandro Besharati

Evaluating Development Effectiveness

Assessing and comparing the impact of education interventions in South Africa

November 2016

A thesis submitted to the Faculty of Commerce, Law and Management, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor of Philosophy.

Abstract

This research is a contribution to the field of development evaluation. Much of the evaluation practice in development and public policy remains weak in scientific validity, and challenged by issues of attribution and comparability of results across different studies. After an in-depth review of the existing literature and an analysis of the current shortfalls and knowledge gaps in programme evaluation, the research proposes a methodological framework that allows for the empirical measurement and comparison of the impact of diverse types of interventions aimed at addressing a specific outcome of interest. The evaluation framework informs decision-making in social-economic development processes, by combing elements of theory-based counterfactual evaluation, multiple-treatment meta-analysis, mixed methods, and participatory approaches. The evaluation framework is tested in South Africa by utilising the proposed package of methods through two case studies presented in this thesis, to generate evidence for policy-makers, programme managers, and investors operating in the education sector. The first is an evaluation of the impact of the corporate social investments of Anglo American Platinum in Limpopo and North West provinces, that utilised geo-spatial features of mining operations to conduct a quasi-experiment. The second is a comparative analysis of major interventions implemented in South Africa to improve learning outcomes in public schools. The education meta-analysis is the first of its kind to be conducted in South Africa, and has revealed many locally-produced impact studies which had not previously been captured by international reviews on school interventions in developing countries. The empirical work conducted in this research confirms existing theories and reveals new insights into the role of the private sector, the proximity of schools to mines, psycho-social and economic factors, learner age and home language, educational material, quantity and quality of teachers, school management, and accountability systems, in affecting education outcomes. The research highlights some of the programmes and policies which have been most effective in South Africa's schooling sector, while cautioning about the contextual factors and methodological design features which influence the effect sizes being reported in the evaluations of development interventions. The research concludes by reflecting on the experiences, data and cost analysis challenges, and the lessons learnt from the application of the proposed evaluation approaches in South Africa's education sector. It discusses the limitations of the framework, and how this can be further refined for future use in other countries, sectors, and development policy contexts.

Acknowledgements

I would like to thank the numerous people and institutions who have accompanied me on this PhD journey. My supervisor and statistics mentor, Michael Muller, the various academics who have provided advice on the contents and structure of my doctorate, Iraj Abedian, Brahm Fleisch, Anne McLennan, Gillian Godsell, Mark Orkin, and the numerous international experts in the field of evaluation methods (many of them referenced in this thesis) whom I have been in contact with, and who provided valuable inputs into my research work. I am grateful for the collaboration of the various South African institutions, stakeholders and researchers in providing crucial data, without which this research would have not been possible. My hard-working research assistants, Khotso Tsotsotso, Jaya Sojen and Mpho Ntlhola, who provided vital support during the field work, data cleaning, and final editing of the thesis. Finally, I would like to acknowledge the patience, understanding, and encouragement shown by my family and friends in this arduous 6-year journey, and the manifold opportunities and blessings I have received from the Almighty God, which has allowed me to be where I am today. Thank you all for being part of my academic and professional growth and supporting me in the path to learning and knowledge contribution.

Table of contents

Abstract 2
Acknowledgements 3
Table of contents 4
List of figures 8
List of tables 10

Chapter 1: Introduction 13

1.1. Background: the macro development discourse 13
1.2. Problem statement 18
1.3. Research questions 21
1.4. Purpose statement 22
1.5. Piloting the evaluation framework in South Africa's education sector 22
1.6. Outline of the thesis 24

Chapter 2: Literature review, theoretical and conceptual framework:
Models, methods and approaches for evaluating development effectiveness 26

2.1. Introduction 26
2.2. The role of evaluation in development 27
2.3. Programme theory, logic models, and results-based frameworks 29
2.4. Typologies of development evaluation 33
2.5. Validity and data quality 36
2.6. Monitoring and the role of indicators 37
2.7. Inputs and resource tracking 39
2.8. Limitations of development agency evaluations 42
2.9. Measuring impact and the challenge of attribution 43
2.10. Non-experimental and econometric approaches 46
2.11. The micro-macro paradox 48
2.12. Randomised control trials 49
2.13. Quasi-experimental designs 53
2.13.1. Matching 54
2.13.2. Cut-off point design 55
2.13.3. Instrumental and control variables 56
2.13.4. Double difference 57
2.13.5. QEDs versus RCTs 58
2.14. Caveats with measuring impact 60
2.15. Unravelling the 'black box' 61
2.16. Mixed methods and participatory approaches 64
2.17. Comparative evaluations and systematic reviews 67
2.18. Comparing effectiveness using meta-analysis 70
2.18.1. Traditional meta-analysis 70

2.18.2. Comparative meta-analysis 71
2.18.3. Study quality and risk of bias 73
2.18.4. Standardising effect sizes 74
2.18.5. Data extraction and confidence intervals 76
2.18.6. Analysing results of MA 77
2.18.7. Exploring heterogeneity through moderators 78
2.19. Cost-effectiveness analysis 79
2.20. Conclusion: developing a conceptual and methodological framework 82
2.20.1. Purpose and requirements 82
2.20.2. Elements of the proposed evaluation framework 83
2.20.3. An organic process of development, piloting and refinement 91

Chapter 3: Contextual overview :

South Africa's education sector and interventions to improve learning outcomes 93
3.1 Introduction 93
3.2 Historical and legislative context 95
3.3 Infrastructure, equity and access 97
3.4 Curriculum frameworks 99
3.5 National learner assessments 100
3.6 The 'crisis' of quality 104
3.7 Funding education development in South Africa 110
3.7.1 Government Spending 110
3.7.2 Parental Spending 113
3.7.3 Private sector investments 114
3.7.4 Other partners in education development 119
3.8 Interventions to improve learning outcomes 121
3.8.1 Family, community and society 124
3.8.2 Early childhood development 126
3.8.3 Language Policy 127
3.8.4 Learning and teaching support material (LTSM) 129
3.8.5 The central role of teachers 131
3.8.6 School management and principals 140
3.8.7 Systemic reforms and sub-national policies 142
3.8.8 Learner-targeted initiatives 144
3.8.9 Integrated school development models 147
3.9 A theory of change for South Africa's education sector 150

Chapter 4: Case Study:

Evaluating the impact of Amplats education programmes in Limpopo and North West 155

4.1 Introduction 155
4.2 Contextual background of the case study 157
4.2.1 The mining sector in South Africa's development 157
4.2.2 The state of the schooling sector in Limpopo and North-West 160
4.3 The Amplats education programme, its theory and evaluation 167
4.4 Case study design and methodology 172

4.4.1 Guiding questions 172
4.4.2 Qualitative field work 172
4.4.3 Data preparation 173
4.4.4 Impact evaluation methods and techniques 182
4.5 Evaluation findings 183
4.5.1 Results from the qualitative field work 183
4.5.2 Results from the econometric analysis 187
4.5.3 Results from the quasi-experimental evaluation 192
4.6 Conclusions 208
4.6.1 Re-affirming the micro-macro evaluation paradox 209
4.6.2 Bottom line for the social investor 210
4.6.3 Participation and policy 211
4.6.4 Follow-up work on the evaluation framework 213

Chapter 5: Comparative meta-analysis of interventions to improve learners achievement in South African schools 214

5. 1 Introduction 214
5.2 Background 215
5.2.1 Challenges of South Africa's education sector 215
5.2.2 The interventions: policies, programmes and investments 217
5.2.3 Growth of evidence-based education in South Africa 221
5.3 Objective 223
5.4 Conceptual Framework 223
5.4.1 Education evaluations in South Africa 223
5.4.2 Standardising impact measures across studies 226
5.4.3 Multiple treatment meta-analysis (MTMA) 229
5.4.4 Integration of cost-effectiveness analysis 232
5.5 Systematic review protocol 233
5.5.1 Inclusion criteria 233
5.5.2 Identification strategy 233
5.5.3 Appraising quality and bias in the studies 237
5.5.4 Data extraction and follow-up 239
5.6 Qualitative results of the systematic review 241
5.6.1 Overview of studies and interventions included in the review 241
5.6.2 Qualitative assessment of the evaluations included in the review 244
5.6.3 Problematic studies excluded from the meta-analysis 270
5.6.4. Adjusting standard errors to cluster sampling 272
5.7 Results of the statistical meta-analysis 274
5.7.1 Effect sizes at face-value 274
5.7.2 Interpreting results in context 276
5.7.3 Exploring heterogeneity through moderator analysis 279
5.7.4 Meta-regression and meta-ANOVA 282
5.7.5 Sub-group analysis 288
5.8 Failure of cost-effectiveness comparison 304
5.9 Participation and peer reviw 307
5.10 Concluding summary 308

Chapter 6: Conclusion **312**

6.1 Evaluation framework emerging from the research ***312***
6.2 Key findings for South African education policy ***314***
6.3 Reflections and learning from the application the evaluation framework ***316***
6.3.1 Measuring impact 317
6.3.2 Comparing effectiveness 320
6.3.3 Participation, peer-review, policy and practice 321
6.3.4 Challenges and limitations 322
6.3.5 Future application to other contexts and policy sectors 324

Reference list ***326***

List of figures

Figure 1.	An example of a Theory of Change.	30
Figure 2.	Illustration of a results chain.	31
Figure 3.	Diagram representating the types of evaluations to be utilised at different levels of the results chain.	35
Figure 4.	Transformation of funds into results in the public and development sector.	40
Figure 5.	Annotation of a simple pretest-posttest randomised experiment.	51
Figure 6.	Graphical representation of the regression discontinuity method.	55
Figure 7.	Graphical illustration of pre-post difference-in-difference impact evaluation.	58
Figure 8.	Matrix of experimental and quasi-experimental methods in relationship to internal and external validity.	59
Figure 9.	Diagram of a black-box.	62
Figure 10.	Evidence hierarchy pyramid	68
Figure 11.	Diagram representing direct and indirect comparison.	71
Figure 12.	Illustration of an indirect treatment comparison.	72
Figure 13.	Graphical representation of effect size *d*.	75
Figure 14.	Forest plots (without pooled effect size).	77
Figure 15.	Graphical representation of network meta-analysis.	89
Figure 16.	Mean reading and maths scores (SACMEQ III).	106
Figure 17.	Grade 12 pass rates and Bachelor passes.	107
Figure 18.	NSC results for math and science subjects.	108
Figure 19.	NSC pass rates by population group.	109
Figure 20.	Percentage of SA national budget allocations by sector.	110
Figure 21.	Percentage of GDP spent on education.	111

Figure 22.	Corporate social investments versus foreign aid to SA.	116
Figure 23.	Theory of change for the operation of South Africa's education sector.	152
Figure 24.	Locations of platinum deposits in Northern South Africa.	158
Figure 25.	Overall and Bachelor pass rates for NSC examination in South Africa.	161
Figure 26.	Overall and Bachelor pass rates for NSC examination in Limpopo.	162
Figure 27.	Overall and Bachelor pass rates for NSC examination in North West.	162
Figure 28.	Maths pass rates for NSC examination in North West.	164
Figure 29.	Physics pass rates for NSC examination in North West.	164
Figure 30.	Maths pass rates for NSC examination in Limpopo.	165
Figure 31.	Physics pass rates for NSC examination in Limpopo.	165
Figure 32.	Maths pass rates for NSC examination in Limpopo.	166
Figure 33.	Physics pass rates for NSC examination in Limpopo.	166
Figure 34.	Geometrical representation of Pythagoras formula.	180
Figure 35.	Example of application of RDD to geographic scale within mining context.	193
Figure 36.	Maps of geo-locations of Mogalakwena mine *vis-à-vis* treated and untreated schools.	194
Figure 37.	Positive and negative forces affecting learning outcomes.	203
Figure 38.	Graphical illustration of standardised mean difference.	227
Figure 39.	Illustration of forest plot in a comparative meta-analysis.	230
Figure 40.	Graphical illustration of multiple-treatment meta- analysis.	231
Figure 41.	Funnel plot of effect sizes by standard errors.	271
Figure 42.	Regression of Hedge's *g* on grade.	284

List of tables

Table 1.	Classical Experimental Model	85
Table 2.	National Expenditure Per Quintile	113
Table 3.	Largest Corporate Education Investors in SA	118
Table 4.	NSES Education Production Function for Learning Gains	122
Table 5.	Teacher Qualifications in South Africa in 1988	134
Table 6.	List of Major Integrated Whole School Development Programmes Implemented in South Africa	149
Table 7.	Intervention Variables used in the Study	175
Table 8.	Outcome Variables used in the Study	176
Table 9.	Explanatory Variables Used in the Study	177
Table 10.	Econometric Analysis on the Impact of Various Variables on Maths Results: Regression on Maths Passes (above 30% score)	189
Table 11.	Econometric Analysis on the Impact of Various Variables on Maths Results: Regression on Maths Excel (above 50% score)	189
Table 12.	Econometric Analysis on the Impact of Various Variables on Physics Results: Regression on Physics passes (above 30% score)	190
Table 13.	Econometric Analysis on the Impact of Various Variables on Physics Results: Regression on Physics excel (above 50% score)	190
Table 14.	Covariates Used for PSM	196
Table 15.	Sample Sizes PSM Proximity to Mines	196
Table 16.	Independent Samples Test on Groups Close to Mine and Far From Mine	198
Table 17.	Effect Size of 'Proximity to Mine Effect'	199
Table 18.	Handicap on Pass Rates Caused by Schools' Proximity to Mining Operations	200
Table 19.	Sample Sizes PSM on Overall Anglo Programme	201
Table 20.	Independent Samples Test on Groups Receiving Support From Anglo and Those That Do Not	202

Table 21.	Effect Size of Overall Anglo Programme	202
Table 22.	Average Results From Pre and Post Test of Learners Attending 2012 Winter Enrichment Programme	204
Table 23.	Percentage of Learners in Treated Schools Attending 2012 Winter Schools	205
Table 24.	Sample Sizes of PSM on 2012 Winter School	206
Table 25.	Independent Samples Test on Results of Schools Attending 2012 Winter Schools Versus Those That Do Not	206
Table 26.	Risk of Bias Rating	239
Table 27.	Summary Statistics Gathered for Studies in the Meta-analysis	239
Table 28.	Studies and Interventions Included in the Systematic Review	242
Table 29.	Studies Excluded from the Statistical Meta-analysis	270
Table 30.	Change in Standard Error Once Adjusted by Sample Clusters	273
Table 31.	Overview of Effect Sizes of SA education Interventions	274
Table 32.	'Rule of thumb' Interpretations of Effect Sizes	276
Table 33.	Normal Annual Learning Gains in American Schooling System	278
Table 34.	Study Coding Table	280
Table 35.	Main Results for Regression Model, Random Effects (MM), z-Distribution, Hedges's g	283
Table 36.	Analysis of Heterogeneity Using Study Design Parameters	285
Table 37.	Analysis of Heterogeneity by School Phase	287
Table 38.	HLM Regression with School Phase (Level 1), Studies/Interventions (Level 2)	288
Table 39.	Interventions To Improve Language Abilities	289
Table 40.	Interventions To Improve Science And Other FET Subjects	290
Table 41.	Impact of Teacher-Based Interventions	291
Table 42.	Impact of LTSM Provision	292

Table 43.	Impact of Whole School Development Programmes	293
Table 44.	Education Programmes Led By Private Sector	295
Table 45.	Impact Of Government Programmes And Academic Experiments	296
Table 46.	Impact Of Education Interventions In Gauteng	298
Table 47.	Impact Of Education Interventions In Western Cape	299
Table 48.	Impact Of Education Interventions In Limpopo	300
Table 49.	Impact Of Education Interventions In Kwazulu Natal	301
Table 50.	Impact Of Education Interventions In Mpumalanga	302
Table 51.	Impact Of Education Interventions In North West	303
Table 52.	Interventions For Which Some Cost Information Was Retrieved	304

Chapter 1: Introduction

1.1. Background: the macro development discourse

Half a century of international development has witnessed the channelling of trillions of dollars into the developing world to address poverty, infrastructure, health, and education challenges (Sachs, 2005; Sen, 1999; Easterly, 2007). Governments, non-governmental organisations, bilateral, and multilateral donors have experimented with policies, programmes and development recipes to improve the conditions of countries and communities throughout the world. The global development enterprise of the second half of the twentieth century was backed up by manifold economic development theories which emerged during this period such as the big push model (Murphy, Shleifer, & Vishny, 1989; Rosenstein-Rodan, 1943), classical growth theory (Hollander, 1998), international dependency theory (Prebisch, 1959; Smith, 1985), and neoclassical free market theory (Blanchard, 1987; Buchanan & Tullock, 1999; Friedman, 2002) which dominated the 1980s aid industry and operation of the Washington-based institutions (i.e. IMF, World Bank, and USAID).

Human development theory (Alkire, 2002; Haq, 1995; Sen, 1999) eventually emerged as the dominant framework animating the operations of the World Bank, the United Nations system, and most bilateral aid agencies and international NGOs (Hicks, 1997). The pinnacle of the application of human development to global policy was the historic adoption of the Millennium Declaration (UN, 2000) by political leaders of the world, which later set the foundation for the UN Millennium Development Goals (MDGs) (UN, 2002). The MDG framework was based on an international development paradigm, which expected developing nations to conduct most of the reforms required to meet Goals 1 to 7, while industrialised countries would provide support to the global enterprise through aid, trade, debt relief and other forms of assistance. The financing model set out by the MDGs was heavily based on a twentieth-century North-South aid paradigm (Besharati, 2013c). In other words, donor countries were called to boost their official development assistance (ODA) to developing countries and to the multilateral agencies (Sachs 2005), who would in turn undertake the projects, policies, and activities of poverty alleviation and social reform, often through local and international NGOs.

An important platform for standard-setting, monitoring and peer-review of donor countries, the OECD's Development Assistance Committee (DAC) took a prominent role in spearheading the aid effectiveness debates through a series of High-Level Forums (HLFs) held in Rome (2003), Paris (2005), Accra (2008), Busan (2011), and Mexico City (2014) which set out the principles for good aid practice and healthy development partnerships between provider countries and recipient countries. Some of the frameworks that emerged out of these meetings such as the Paris Declaration (2005) (OECD, 2008a) and the Busan Global Partnership (2011) (OECD, 2011), were accompanied by indicators and monitoring systems which allowed for periodic tracking of progress on international commitments and the evaluation of the quantity and quality of international development efforts. This encouraged a competitive culture among rich countries to increase aid volumes and improve donor behaviour in favour of developing countries (DIIS, 2011).

The global campaign of the MDGs drew to a close with a mixed bag of successes and failures (UN, 2013). The global community has recently entered a new chapter in the international development undertaking, to eliminate poverty, reduce inequality, promote peace, and safeguard the planet through Agenda 2030 and the new set of 17 Sustainable Development Goals (SDGs).

Education has always been a central pillar of national and global development strategies and closely linked to the achievement of all other goals (UNESCO, 2015). Studies (e.g., Carnevale, Rose, & Cheah, 2011; Macdonald, Barrera-Osorio, Guaqueta, Patrinos, & Porta, 2010) have shown for instance that each year of schooling is associated with 10 to 20% increase in wage earnings. The previous MDG 2 played an important role in ensuring that now 91% of the children of the world have access to primary education (UNDP, 2015), especially girls. In the SDG 4 the challenge has now been raised to achieve 'inclusive and equitable quality education and lifelong learning for all'. The World Education Forum, held in Incheon in May 2015 (World Bank, 2015a), re-affirmed commitment by governments to:

"…transform lives through education, recognizing the important role of education as a main driver of development and in achieving the other proposed SDGs… We reaffirm that education is a public good, a fundamental human right and a basis for guaranteeing the realization of other rights. It is essential for peace, tolerance, human fulfilment and sustainable development. We recognize education as key to achieving full employment and poverty eradication." (Incheon Declaration, 2015, Par. 5)

Policy-makers and researchers (i.e. DBE, 2010; Hanushek & Woessmann, 2012; Motala, Dieltiens, & Sayed, 2012; UNESCO, 2015) have emphasised that simple access to education is no longer enough; improving the quality of education and learning outcomes in developing countries must be given highest priority in order to fully overcome poverty and inequality.

The above global developments have been occurring against an evolving political and economic backdrop. Many countries, which only thirty years ago were war-ravaged and poverty-stricken, have had an exceptional rise and have emerged as new global and regional powers. Mexico, Chile, and Korea (recent additions to the OECD), Brazil, Russia, India, China, and South Africa (the BRICS), the OPEC countries, striving middle-income economies in Africa, Asia, and South America (Schulz, 2010), are playing a growing role in international development (Sidiropolous, Pérez Pineda, Chaturvedi, & Fues, 2015). Some of these countries contribute between 1 and 7 billion dollars a year in development cooperation (Besharati, 2013; UNDESA, 2010), surpassing in some cases the assistance provided by some of the smaller OECD donors. The most crucial national development processes occur also in middle-income countries, which house 5 out of the 7 billion peoples of the world (World Bank, 2015b). While on the global stage, many of these countries have risen to the table of the 20 largest economies of the world (G20), the 'new bottom billion' (73% of the world's poor) reside in these same middle-income countries (Glennie, 2011; Sumner, 2012). India, for instance, while being among the 10 largest economies, has also the same amount of people living below $2 a day poverty line as the whole of Sub-Saharan Africa (Besharati, 2013c). Many of these emerging economies, such as South Africa, Brazil, Colombia, Namibia, in fact, present extremely high rates of income inequality (World Bank, 2016).

Africa has generally also witnessed an unprecedented economic spurt (Mahajan, 2011) with numerous developments driven by endogenous initiatives and a booming natural resource industry (Bello & Manrique, 2011). Private capital is beginning to play a more prominent role in development, through philanthropy, corporate social responsibility, social impact bonds, impact investing, climate change financings, pro-poor innovation, and other important contributions private sector makes to employment, infrastructure, and enterprise development in Africa. At an OECD-DAC meeting in December 2012, USAID representative commented that while foreign aid of US government has been hovering around $30 billion, the American philanthropic sector has been providing over $40 billion in global charity, and American foreign direct investments in the developing countries has reached $3 trillion every year (Besharati, 2013a). To take an African illustration, in the Tete rural province of Mozambique, social investments by foreign coal mining companies have been larger than the ODA provided by bilateral and multilateral donors to that province (Besharati, 2012). Even in the middle-income country context of South Africa, the domestic private sector contributes over 9 billion rands a year in corporate social investments (CSI) (Trialogue, 2015), surpassing by and large the total aid received by South Africa from all its traditional donors (National Treasury, 2015). In the education sector, this contrast is even more acute where CSI and philanthropy by the private donors is ten times larger than that of traditional bilateral and multilateral development agencies (Besharati, 2015). This unique South African case will be further explored in this research.

In short, the development landscape has drastically evolved and now presents an ever-more complex web of very diverse interventions, approaches, modalities, mechanisms, partnerships, and institutions endeavouring to alleviate poverty and improve the living conditions of the peoples on the planet. The new emerging economies, the private sector and its philanthropic foundations, and various forms of innovative development financing, along with the traditional development partners, present a new complex development architecture and a diversity of partners and financial flows which developing countries have to now deal with. Empirical evidence and knowledge generated from the analysis and evaluation of these new players and development finance modalities and mechanisms are crucial in informing global, regional and national development planning processes.

The OECD-DAC monitoring systems and the great bulk of aid effectiveness research and policy focuses on assistance from governments and multilateral institutions, and have neglected large volumes of non-ODA flows that reach every year to the developing world. The Brookings Institution (Kharas & Linn, 2008) argues that private aid flows, together with aid from the non-DAC donors reach up to $60 billion a year and account for more than half of the aid received by developing countries. The Hudson Institute has calculated that global philanthropy to developing countries has reached $50 billion (Hudson Institute Center for Global Prosperity, 2013), making a striking contrast to the World Bank's global contribution of $20 billion each year (Desai & Kharas, 2010). Total global philanthropy amounts to a larger volume of funds than the GDP of more than 60% of the world's economies (World Bank, 2010). In emerging African economies such as South Africa, Nigeria, and Kenya, 15 to 35% of private equity and assets (approximately USD 680 billion) is channelled towards impact investing (Giamporcaro & Dhlamini, 2015). Year after year, ODA flows have been dwarfed when compared to other sources of financing to the developing countries, such as climate change funds, philanthropy, remittances, revenue from domestic taxes, customs, and royalties, and CSI, particularly from the extractive industries in Africa (Greenhill, Prizzon, & Rogerson, 2013; Lundsgaarde, 2013).

There is less information available on these alternative resource flows, as they often by-pass official reporting systems and operate outside the national governments and OECD systems. Desai and Kharas (2008) argue that these unofficial development flows are potentially more responsive, flexible, less tied by political interest, less subject to corruption, with less overheads, and free from public sector inefficiencies. There is however not enough empirical evidence that suggest that such alternative channels of development support are any more or less effective than those from traditional development agencies and governments. Major knowledge gaps and evidence deficits still exist in global development policy, with regard to the role of new private sector players, and the innovative interventions, modalities, and approaches to development. This poses also challenges to the global and national systems with regard to the monitoring, evaluation, and information management of such development undertakings.

Since the Accra (2008) and the Busan (2011) HLFs there has been a major shift of discourse from aid effectiveness to 'development effectiveness'. This became the new buzz word in the international development industry used profligately by all kinds of agencies, but with slightly different interpretations and understandings. The North-South Institute (Kindornay & Morton, 2009) makes a useful attempt to untangle the various meanings behind the concept and to explain the different facets of development effectiveness promoted by the various institutions and the different fora. Kindornay and Morton (2009) explain that development effectiveness has frequently been used by aid agencies, especially multilateral organisations, as a means of assessing the effectiveness of their own policies and programs. Development effectiveness is thus considered from the supply side, in terms of how well an organisation is achieving its stated objectives and goals. This approach has a strong orientation towards organisational performance. AusAID's Office of Development Effectiveness (ODE), for example, has been set up to check the quality and impact of Australia's aid program (AUSAID, 2008). This is closely linked to the need to account to donors and taxpayers, which are pressed by financial crisis and shrinking budgets, and are increasingly concerned with operational efficiency and value for money.

The global South, and in particular Africa, have used the development effectiveness brand to re-define the concept of development, calling on more policy coherence in development (PCD). This approach acknowledges that the achievement of development outcomes is more than just aid. Many parallel forces and factors come into play in a holistic development approach to social and economic transformation. Policy coherence in development calls for the systematic promotion of parallel and reinforcing actions across government departments and agencies in order to achieve increased development results for poor countries (OECD, 2012). The AU-NEPAD together with UNDP in the lead up to Busan HLF4 established the multi-stakeholder African Platform for Development Effectiveness (AP-Dev) which produced the African Consensus on Development Effectiveness (2011). The document reflected the new paradigm, advocated by Africa, in which the development community was urged to break the cycle of aid dependency and to look at new sources of development finance, including remittances, trade and private sector development, foreign and local investment, and most importantly promote domestic resource mobilisation.

A more popular understanding of development effectiveness draws on the notion of results and the contribution that different types of interventions make to development outcomes. Such an approach tracks progress through indicators and emphasises impact on the lives of people. The UNDP has for example published a number of Assessment of Development Results (ADR) and Development Effectiveness Reports (DER), where a useful distinction is made between organisational effectiveness, linked to achievement of lower-levels results (activities and outputs), and development effectiveness,

which measures the change in the condition of beneficiaries (outcomes and impact) brought by the interventions.

The 2007 Development Effectiveness Report of the International Fund for Agricultural Development (IFAD) suggests that development effectiveness depends on the collective and coordinated actions of a range of national and international actors, therefore the impact and accountability needs to also be shared among various stakeholders. The World Bank's Annual Review of Development Effectiveness (ARDE, 2009) looks at its organisational performance, but it makes clear that it cannot establish direct link between its projects and the country's macro-economic performance. In 2002, DFID commissioned an independent review (Flint, Cameron, Henderson, Jones, & Ticehurst, 2002) of its organisational and development effectiveness. The report was candid in concluding that it is virtually impossible to illustrate a direct link between DFID's activities and the progress towards the MDGs, despite various admirable attempts. The overall challenge is that the higher one looks in the results chain the more difficult it is to establish causality and attribution between the intervention and the development change (Johnson & Lamdany, 2005).

In the current development effectiveness debates, evaluation assumes a central role in supporting accountability, evidence for policy-making, learning and improvement of systems, and the maximising of impact of development endeavours in favour of the poorest and most marginalised groups. Many challenges and knowledge gaps nevertheless remain with regard to how to address problems of attribution and comparability of results, when evaluating development interventions. This global policy background sets the basis of the micro-level enquiry that will unfold in this research.

1.2. Problem statement

Notwithstanding large financial investments, broad-based political will, and vast technical expertise accumulated over the decades, the global development enterprise did not achieve the improvement in the social-economic welfare for the vast majority of poor in the world (Collier, 2008; Dietrich & Wright, 2012; Sachs, 2005; United Nations, 2013). Development programmes and policies have grown ever more complex, elaborate, and expensive but are still ineffective in tackling the challenges faced by developing countries (Alesina & Dollar, 2000; Easterly, 2007; Minoiu & Reddy, 2010; Riddell, 2008; Rondinelli, 2013).

Demands on accountability are increasing from citizens in recipient countries who want to see tangible results from development programmes and public services (Bovens, Goodin, & Schillemans,

2014; Tanzi, 2016; Wolfensohn, 2005). Philanthropists and taxpayers in donor countries also want to see their contributions spent more effectively and wisely (Quigley, 2015; Wiesner, 2005). Decision-makers in the development sector are constantly faced with many options and choices, such as where to invest resources, which interventions to favour, which institutions to engage with, and which approaches to take in order to achieve bigger impact on development outcomes. When interventions are not working, it is important to also assess why? Where do the problems emerge from, and what can be done to rectify the system? Development institutions are thus constantly learning, improving, and evolving (Donaldson, 2005; Fowler, 2013; Meinzen-Dick, 2004). In such settings, evaluation is a critical tool to help answer questions and provide empirical evidence for policy-makers and managers who regularly make decisions on development processes that affect the lives of millions of people.

Evaluation in development management and public policy has been acquiring a growing importance and becoming increasingly more complex. Since the last two decades many development agencies started utilising results-based frameworks (Kusek & Rist, 2004; Lahey & Nielsen, 2013; Watson, Broemeling, Reid, & Black, 2014); however, the focus of most evaluative processes is on assessing implementation and organisational efficiency (Cummings & Worley, 2014; Kindornay & Morton, 2009a). This has resulted in a focus on monitoring and reporting of activities and outputs, which are easier for managers to control, and on the volumes of financial inputs, which are of higher interest to donors (Fowler, 2013). Agency-level impact evaluations tend to rely heavily on qualitative methods, such as observations and interviews with field stakeholders. However, these types of studies often run the risk of being very subjective, compromising independence, lacking scientific rigour, and reporting more the successes rather the failures of programmes (Riddell & Kruse, 1997; Royse, Thyer, & Padgett, 2015). Often because of the lack of baseline data, or because of limited budgets for monitoring and evaluation, even simple 'before and after' outcome assessments are not conducted (Bamberger, Rao, & Woolcock, 2010).

In the development economics literature (e.g., Addison, Mavrotas, & McGillivray, 2005; Burnside & Dollar, 2000; Guillaumont & Chauvet, 2001; Neumayer, 2003; Rasella, Aquino, Santos, Paes-Sousa, & Barreto, 2013; Roodman, 2008) the most popular methods used to evaluate the impact of macro-level development policies are different forms of multivariate regression. These econometric models often present very diverse results and are highly contested and criticised for methodological problems (Clemens, Radelet, & Bhavnani, 2004; Deaton, 2009; Doucouliagos & Paldam, 2009; Kmenta & Ramsey, 2014; Moss, Pettersson, & Van de Walle, 2006,). Mosley (1986), Boone (1996), Picciotto (2006) and Ndikumana (2012) have furthermore found that a 'micro-macro paradox' exists in development evaluation where the aggregate of micro-level programme evaluations, mostly reporting

positive results, is not consistent with the macro-level econometric evaluations, which mostly report negative trends.

One of the paramount challenges that development evaluators face is establishing causality and attribution (White & Phillips, 2012) of the impact in relation to the development interventions being evaluated. When assessing the effectiveness of development programmes, the higher one goes in the results chain, the more difficult it is to attribute a specific outcome to the intervention of a particular agency. In fact, the higher-level development outcomes are usually affected by a multitude of agencies, as well as external and internal forces, and diverse social, economic, and environmental factors (Johnson & Lamdany, 2005). In order to tackle this problem, rigorous impact evaluations utilise counterfactual techniques (Gertler, Martinez, Premand, Rawlings, & Vermeersch, 2011) to isolate the effect of interventions from other factors operating in the environment.

Some of the leading development economists (Banerjee, Duflo, Glennerster, & Kinnan, 2015; Kremer & Glennerster, 2011; Glewwe, 2012) have tried to address the above challenges through the use of randomised experiments in the evaluation of social programmes. Such randomised control trials (RCTs) are, in fact, considered by many the 'gold standard' for impact evaluation. Randomised control trials however take extensive time, are expensive, face ethical challenges and in most cases provide results which are very context-specific (Bonell, Hargreaves, Cousens, Ross, Hayes, Petticrew, & Kirkwood, 2011; Sanson-Fisher, Bonevski, Green, & D'Este, 2007; Trochim & Donnelly, 2001). The biggest challenge when using randomised experiments in the social sciences is that these are not always practical to implement in real-life setting and therefore are rarely used in dynamic policy environments where interventions are mostly evaluated ex-post. Experimental methods are often also criticised for encouraging a 'black box' (Bickman, 2000; Chen, 2005; Salter & Kothari, 2014) and 'a-theoretical' approach (Mouton, Wildschut, Richter, & Pocock, 2013) to evaluation which does not provide the how and the why (White, 2013) certain interventions work or not.

Although various degrees of methodological quality and scientific rigour are applied in development evaluations, different academics, bilateral and multilateral organisations, corporate investors, NGOs and government agencies, use different paradigms, criteria, standards, approaches, and methods in the evaluation of their development programming (Donaldson, 2005, Patton, 2011). The lack of a common framework to evaluate the effectiveness of development interventions makes it difficult to empirically compare the findings of different evaluations, and thus to compare also the effectiveness of different institutions and different programmes that endeavour to achieve the same developmental objective.

As evaluation of development programming grows in demand, issues of scientific rigour, causality, attribution, and comparability of results across different studies become increasingly important to policy-makers and managers who require ready evidence when taking decisions that affect public affairs and the development of communities.

1.3. Research questions

Overarching research question

What is an appropriate framework for the evaluation of development programmes that can be useful, practical and engaging for development managers and stakeholders; while also scientifically rigorous in the generation of evidence for policy-making?

Research sub-questions

1. How do we empirically measure the impact of development programmes?
    a) How much can the improvement in the condition of beneficiaries be attributed to the intervention of a specific agency? How do we establish causality?
    b) Considering the ethical, financial, and practical constraints of RCTs, are there alternative ways to conduct counterfactual evaluation in development policy settings?
    c) What are the exogenous and endogenous factors that influence the magnitude of the impact caused by a development intervention?

2. How can the effectiveness of different interventions be compared through a common framework?
    a) How can the results from different evaluations and studies (using different methods and scales) be standardised and compared against one another?
    b) How can managers and policy-makers assess what is the most effective intervention to undertake in a particular development context?

1.4. Purpose statement

The research undertaken here presents an analysis of the development evaluation literature to propose a flexible methodological framework that empirically measures and compares the effectiveness of different types of development interventions. It explores approaches and instruments to gather scientific evidence on development processes and assess which interventions produce the most impact when endeavouring to achieve a specific development outcome. The framework endeavours to bring together elements of theory-based quasi-experimental impact evaluation, systematic review and multiple treatment meta-analysis (MTMA), mixed methods, and participatory approaches. The research will thus make a contribution to evaluation practice, development management, public policy, and the current debates around development effectiveness.

1.5. Piloting the evaluation framework in South Africa's education sector

As the conceptual and methodological framework emerges from the review of the literature, the following research endeavours to pilot the proposed set of methods and approaches in a real-life public policy setting, relevant to Africa and most developing countries. The practical application of the evaluation framework is thus illustrated through empirical studies undertaken in South Africa's education sector.

Education is one of the core pillars of the human development paradigm (Haq, 1995; Sen, 1999) and is the bedrock of development of individuals, communities and countries. It provides the knowledge, skills, values required to develop capabilities (Sen, 1999) to live healthy, prosperous, creative, free, and fulfilled lives. The quality of primary education is central in laying the foundations for higher learning, professional development, and generating employment required to propel forward economies. It is thus one of the key factors to address poverty and inequality, which are some of the endemic problems that currently affect the African continent (Adams, 2003; Bhorat, 2008; Fosu, 2009).

Historically in South Africa, education has always played a critical role in political, economic, and social transformation (Bloch, 2009; De Clerk, 2002; Fiske & Ladd, 2004; Fleisch, 2008). In the words of former Deputy-President of South Africa, Kgalema Motlanthe:

"Education is without a doubt a lever to uplift individuals, their families and society at large. Nowhere is this more true than in South Africa, where education should serve as a weapon against the scourge of poverty among our people." (Nevondwe & Matotoka, 2013).

Education therefore features very high in the political and economic priorities of South Africa (i.e. NPC, 2013). A large range of development players – national and provincial government agencies, corporations and foundations, non-governmental organisations, and foreign donors, are all engaged in the sector, investing substantial resources and implementing a range of elaborate school development models, policies, and programmes to address the common imperative of improving the educational deficiencies of the previously disadvantaged learners. South Africa presents also an interesting case on private sector engagement in education, which can offer useful insights into the current global debates on alternative sources of financing for development (Besharati, 2015; Greenhill et al., 2013).

Both national and international assessments (i.e. Annual National Assessment (ANA); Progress in International Reading Literacy Study (PIRLS); Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ)) show that quality of South Africa's education system is still pitifully low (OECD, 2015; Schwab, 2011), below the standards of similar middle-income countries, and even of poorer countries within the African region. The large investments in the sector, amounting every year to 5-6% of the country's GDP (OECD, 2008b), and the variety of interventions to improve the public education system, have unfortunately not translated into the desired results hoped by policy-makers and development managers. Notwithstanding a vast number of education studies conducted post-1994, there is still limited information on the successes and failures of school interventions, and the lessons that can be extracted therefrom for public policy (Bloch, 2009; Motala & Pampallis, 2005; Sayed, Kanjee, & Nkomo, 2013; Taylor, Muller, & Vinjevold, 2003). Most of the international reviews on education programming in developing countries (i.e., McEwan, 2014, Glewwe et al., 2014, Krishnaratne et al., 2013) contain only a limited number of education evaluations produced in South Africa (Evans & Popova, 2015); while the latest most comprehensive South African education review (Mouton et al., 2014) does not contain a systematic and statistical approach for the comparison of interventions.

The evaluation framework developed in this research, is thus utilised to gather empirical evidence to answer important questions for South Africa's education policy. The methods and instruments proposed in this research are illustrated in the evaluation of development programmes undertaken in South Africa's schooling system. The framework is used to empirically compare the impact of different types of interventions of both public and private investors, traditional and new development partners. It helps identify some of the most effective programmes and policies that can be considered for replication, scaling, and learning. It explores the key features, factors, and contexts that contribute to the

success of interventions aimed at improving the overall quality and delivery of South Africa's education system.

The research applies the methodological framework to two specific studies presented in this thesis. The first is the evaluation of the impact of the social investments of mining company Anglo American Platinum towards local communities in Limpopo and North West provinces. The second is a comparative meta-analysis of some of the most prominent programmes implemented over the past two decades aimed at improving learning outcomes in South African public schools.

1.6. Outline of the thesis

Chapter 1 provides a macro-level background on the international development landscape, its historic evolution, and the current knowledge gaps in the development effectiveness discourse. It has unpacked the problems and challenges around development evaluation, and set out the questions and purpose of the research, which is to propose a conceptual framework for the empirical assessment and comparison of the effectiveness of different types of development interventions. It has also explained that the evaluation framework and the set of methodologies proposed in this research will be illustrated in South Africa's education sector to provide evidence for policy-makers and investors in this field.

Chapter 2 conducts a thorough literature review of the main tenets, models, approaches, techniques, and methods used in development evaluation, providing a critical analysis of their respective strengths, weaknesses, and utilisations. This will lay the theoretical foundations for the evaluation model, conceptual frameowkr, study design, and methodologies used in chapter 4 and 5 of this thesis. The research proposes a development evaluation framework that combines elements of theory-based evaluation, systematic review, multiple treatment meta-analysis, quasi-experimental evaluation, cost-effectivenss analysis, mixed methods, and stakeholder participation.

Chapter 3 provides the context and policy setting in which the evaluation framework will be applied through the empirical case studies in the subsequent two chapters. This chapter will provide an overview of South Africa's education sector, the historical socio-economic and political challenges, the institutional and legislative context, the major education investments by both public and private actors, the strength and weaknesses of the national learning assessment instruments, and the diverse types of programmes and policies implemented by different institutions to improve the poor-quality learning

outcomes in public schools. This will include also a discussion around the theory of change, which underpins the overall functioning of South Africa's education sector.

Chapter 4 utilises the evaluation framework to conduct an original impact study on the effectiveness of the Anglo American Platinum education programme and its individual components. This forms part of the evaluation of the company's social investments made between 2009 and 2012 in public schools in Limpopo and North West provinces of South Africa. Through the use of econometric, geo-spatial quasi-experimental methods as well as qualitative fieldwork, the chapter confirms previous education theories, reveals surprising findings, and offers new evidence with regard to the effects that mining operations, development interventions, and social-economic factors have on learning results of surrounding schools and affected communities.

Chapter 5 further expands the application of the evaluation framework by conducting a comparative meta-analysis and cost-effectiveness analysis of some of the major programmes and policies to improve learner achievement in South African public schools. The chapter offers to the international education literature a systematic review and meta-analysis of 28 experimental and quasi-experimental evaluations conducted in South Africa's education sector in the past 15 years, revealing which programmes had the biggest impact on language, mathematics, and science outcomes. It will also look at some of the contextual factors, methodological and designs features that influence the effect sizes reported for each of the education interventions.

Chapter 6 concludes the thesis by summarising the key findings from the empirical work and the new knowledge generated through the research process. It reflects on the methodological framework developed in the research and on the experiences and lessons learned in its application to South Africa's education sector. It will highlight the requirements, merits, strength, weaknesses, context, usefulness, and limitations of the evaluation framework and how it assisted in answering the initial research questions and the problems currently existing in development evaluation community. It will discuss the challenges encountered in conducting cost-effectiveness analysis and suggest potential follow-up work around the framework, proposing refinements and improvements in future applications to other sectors, countries and development policy contexts.

Chapter 2: Literature review, theoretical and conceptual framework: Models, methods and approaches for evaluating development effectiveness

2.1. Introduction

Macro-development debates (Collier, 2008; Desai & Kharas, 2008; Easterly, 2007; Greenhill et al., 2013; Kindornay & Morton, 2009; Mahajan, 2011; Moyo, 2009; Sachs, 2005; Sen, 1999) have been unpacked in the previous chapter and also some of the knowledge gaps that currently pose a challenge to analysts, managers, and policy-makers in international development. In a complex development landscape with a diversity of players, programmes, partnerships, and implementation modalities, development practitioners are faced with many choices, and therefore need to have at hand empirical instruments that allow them the opportunity to assess problems in the public system, identify interventions and approaches that have proven more effective, and allocate resources more efficiently for greater impact on beneficiaries of their programmes (Wiesner, 2005). Accountability, ongoing learning, and improvement are at the heart of all endeavours of all development institutions - local, national, regional, or international (Donaldson, 2005; Meinzen-Dick, 2004).

In such settings, evaluation becomes a critical tool to address questions and provide empirical evidence for decisions on development processes that affect the lives of millions of people (Wolfensohn, 2005). The practice of development evaluation has evolved over the decades, but many challenges still persist such as addressing causality and attribution of development results (Johnson & Lamdany, 2005; White, 2005) and conducting evaluations of a certain standard and scientific rigor (Bamberger et al., 2010; Gertleret al., 2011). Academia, bilateral and multilateral development agencies, philanthropies, NGOs, and governments, use different paradigms, criteria, approaches, and methods in the evaluation of development programmes and policies (Donaldson, 2005). The lack of a common framework to evaluate the effectiveness of development interventions makes it difficult to compare the findings of different evaluations, and thus to compare the effectiveness of different types of interventions (White, 2005).

At the heart of the following research is the development of an evaluation framework that allows for the empirical measurement of the effectiveness of different types of development interventions. This

necessitates that the methodological approach be flexible and adaptable to different sectors and institutions, and useful in the processes of public policy, resource allocation, and development management. The evaluation framework thus intends to provide the conceptual and scientific instruments needed to produce evidence that can assist policy-makers in identifying interventions that produce the best results and best value in the achievement of outcomes within a specific developmental context.

An in-depth review of the evaluation literature and of the evaluation frameworks, methods, and practices that dominate the development sector is presented in this chapter. It will reflect on some of the prominent tenets, models, approaches, and techniques used in the evaluation of socio-economic development programmes and public policies. A critical analysis of some of the most common development evaluation frameworks, illustrating the questions they try to address, the criteria they use, their respective strength and weaknesses, the methods they exploit, their inherent limitations, and the settings in which they are normally undertaken, is presented. It will delve in further detail on the subject of impact evaluation and approaches to comparative analysis of development interventions, which constitute the backbone of the conceptual framework proposed in this study. The methodological review will draw out experiences and insights from the academic and policy literature that can contribute to the development of an appropriate evaluation framework to be utilised in this research. This chapter will eventually lay the theoretical foundations for the evaluation model, design, approach, and methodologies used in chapters 4 and 5 of this thesis, where the framework will be piloted to conduct empirical work in South Africa and inform education policy, programming, and social investments.

2.2. The role of evaluation in development

Five decades of development endeavours, dedicated investments, and experimentation with diverse programmes and policies by governments, NGOs, and development agencies have sadly not led to the results that were hoped for (Collier, 2008; Easterly, 2007; Moyo, 2009; Sachs, 2005; Sen, 1999). Often important policy decisions and resource allocations in development management are made based on limited experiences, personal opinions, assumptions, and ideology, rather than on empirical evidence, which research and evaluation could have provided (Davies, 2012).

Over time, assessing the results of development intervention has assumed growing importance among development practitioners and policy-makers. The demand for evaluation comes from donors and taxpayers, who want to see if their money is being spent well, as well as from recipients of aid, who are keen to know what concrete impact these interventions are having on their daily lives (Wolfensohn, 2005). Evaluation, therefore, is closely linked to the concept of accountability and often tied to budget

allocations and fund disbursements (Wiesner, 2005). Endogenous evaluation can also help institutions develop their effectiveness in what they are doing (Donaldson, 2005). When undertaken internally, without threat of external criticism, evaluation can encourage reflection and a culture of learning which leads to enhanced efficiency in development operations, ironing out problems, and improving the way things are done (Meinzen-Dick, 2004). It helps to avoid that which does not work and adjust policies and programmes to maximise results; therefore, it is critical for the improvement of development interventions and practices.

The core purpose of evaluation is therefore to gather evidence that will assist policy-makers and managers to make more informed decisions (Davies, 2012). It can be used to diagnose a situation, the size, the dynamics, and the roots of a problem, and can help in the design of an intervention and the development of the theory of change (TOC) to address a specific policy area. It checks whether activities are on track, assists in understanding divergence between planned and actual performance, and identify emerging issues. It provides evidence for resource-allocation and for choosing between different alternative and competing strategies (Lindahl & Catterson, 2005). It can support reform, innovation and improvement in public policy and institutional structures. It is often done at the end of an intervention to assess if objectives have been met or not, and why? It endeavours to gather lessons and good practices that will provide recommendations for future programming (Leeuw & Cooksy, 2005) and potential replication and scale-up of intervention models. Evaluation, therefore, performs a central function in development management.

Evaluation is about analysing performance and assessing if results have been achieved. The Centers for Disease Control and Prevention (CDC, 1999) defines evaluation as an analytical effort to answer specific questions about performance of programmes. Scriven (1991) defines it as "the process of determining the merit, worth, or value of something, or the product of that process". Rossi, Lipsey, and Freeman (2004, p. 20) defines evaluation as a "family of research methods which seeks to systematically investigate the effectiveness of social interventions in ways that improve social conditions". Finally, Morra-Imas and Rist (2009, p. 108) define evaluation as "an assessment of planned, ongoing or completed intervention to determine its relevance, efficiency, effectiveness, impact and sustainability. The intent is to incorporate lessons learned into the decision-making process."

In essence, evaluation follows the same principles, approaches, and methods, and faces the same challenges of social research in general. It is, however, a practical application of the science and methods of research to the field of management and public policy. Therefore, critical to the evaluation process is what precise questions are being asked. What exactly is being evaluated and why is the evaluation being conducted? For which target audience is it intended? How are interventions being appraised and judged

and with what criteria? What types of evaluation can be undertaken and what methods, approaches, and design frameworks can be used?

2.3. Programme theory, logic models, and results-based frameworks

Before continuing the discussion around evaluation it is important to reflect on some of the major theories and frameworks that govern the disciplines of programme management, development planning, and public policy. Central to implementing any evaluation is unpacking first what the intervention is all about and how it works. An intervention is hereby defined as any development project, programme, strategy, policy, initiative, or activity aimed at improving the socio-economic conditions of people. These can be implemented at individual, household, community, national, sub-national, regional, or international level.

Every intervention is built on a particular *theory of change* or *program theory* (Chen, 1990; Funnell & Rogers, 2011; Levison-Johnson, Dewey, & Wandersman, 2009; Weiss, 1997). Scriven (1976) had described this as *modus operandi* and modern academicians and practitioners have referred to it in different ways – impact theory, theory of change, logic model, results framework, outcome mapping – but effectively they all refer to the same concept (White & Phillips, 2012). Bickman (1987) and Torvatn (1998) describe it as a *model* of how a programme is supposed to work and Chen and Rossi (1983) have explained how theory helps understand social phenomena. Funnell and Rogers (2011, p. 36) define program theory as "an explicit theory of how an intervention contributes to a set of specific outcomes through a series of intermediate results". It is the theory of how change occurs and how through specific actions certain improvements are made in a specific situation. In simple terms it is the *story* of how change happens in a community, with and as a result of a particular development intervention.

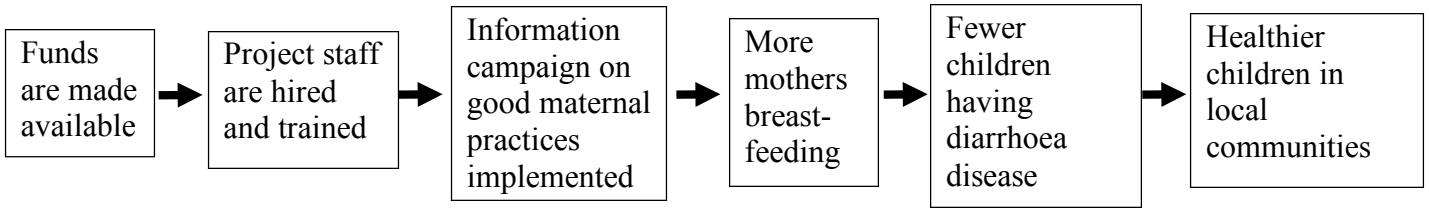
Often program theory is divided into two major parts: one refers to the theory of how the interventions are implemented (theory of action), and the other refers to the changes which occur in the society/beneficiaries as a result of the intervention (theory of change). Weiss (1997) distinguished between *implementation theory*, the chronological steps of a project activity, and *program theory*, the psyco-social transformations that occur in a community receiving an intervention. Donaldson (2007) refers to these two elements as *process theory* and *impact theory*'.

For any one development sector or any intervention there could be multiple theories that are put forth, which could even be at times conflicting with one another. Program theories can be explicit, but most often they are implicit in the conceptual frameworks of institutions and individuals that design and implement development interventions. Programme theories are, in fact, based often on the experiences,

opinions, assumptions, world-views, and ideologies of the program managers and policy-makers who develop them. Strong program theories are rooted in evidence, prior research, and other social science theory (Chen & Rossi, 1987; Funnell & Rogers, 2011), or *theory with the big T* (du Toit & Mouton, 2013). Theories are developed not only through reviewing previous literature on the subject but also through expert and stakeholder interviews and other participatory techniques (see section 2.16).

A key characteristic of intervention theory is that it operates under the principles of causality (Weiss, 1997). It follows a cause and effect chain of events organised in a logical flow. Hence, in the humanitarian and development industry this is often referred to by practitioners as *logic model*. Logic chains, however, are not always linear, but could branch out and have complex structures (Funnell & Rogers, 2011; Patton, 2011). Typically, evaluation experts and development managers like to depict logic models for interventions through graphs, charts, and diagrams using boxes and arrows (White, 2013). An example of a simple Theory of Change (TOC) for a health intervention is illustrated in Figure 1.

Figure. 1. An example of a Theory of Change.



Source: Author's own compilation

Lipsey (1993) describes it as a proposition about what goes on in the 'black box' (see section 2.15) of development interventions. A TOC, represented in a logic model, is in fact, nothing but a "causal hypothesis waiting to be tested" (du Toit & Mouton, 2013, p. 166).

Program theory affects all aspects of the project cycle and is essential for conducting all development management functions, from design, monitoring, implementation, and evaluation, properly (Funnell & Rogers, 2011). Often interventions fail simply because they have been designed with a poor logical model, or the TOC was not based on real evidence (Davies, 2012; White, 2013). Conducting an analysis of the TOC may also reveal risks, assumptions, and other social-economic, institutional, and environmental forces, endogenous and exogenous factors, which affect the development outcomes and therefore can be utilised as confounding variables during impact evaluation processes (See more in section 2.10 and 2.13). It may be concluded that to conduct proper evaluation of any programme or

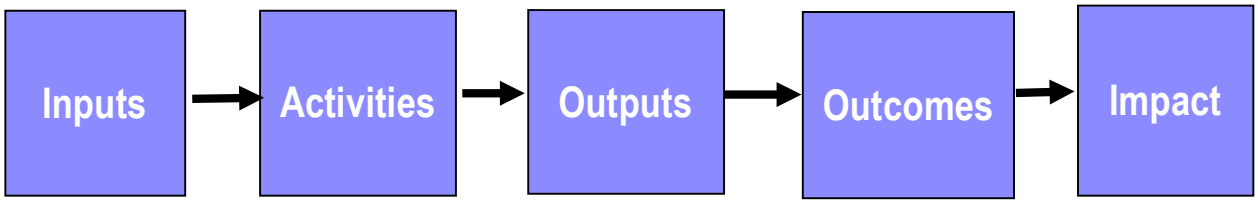
policy, the intervention logic needs to be clearly defined and framed by managers, evaluators, and other programme stakeholders involved.

Program theory is the academic foundation of the practice of results-based management (RBM) which is the most popular framework used today in the development industry. Results-based management is a management approach which emphasises development results in planning, implementation, monitoring, learning, and reporting (UNDP, 2002). Its purpose is to assist organisational learning, improve performance, and enhance accountability and transparency (UNDG, 2011). It assists managers by providing controls and clarity when trouble-shooting (Kusek & Rist, 2004). It provides support to decision-making processes at various stages of the project (CIDA, 2008).

The origins of the results orientation can be traced back to Drucker (1954) and his 'management by objectives' and Suchman's (1967) 'chain of objectives'. A key milestone was the logical framework approach (LFA), which Rosenberg & Posner (1979) had developed for USAID. This was an instrument borrowed from the engineering field and applied very effectively to development planning. By the 1990s RBM became the standard in development management framework used by almost all governments, NGOs, and international organisations, including the World Bank and the United Nations system. The Millennium Development Goals (MDGs) were underpinned by the RBM framework, and even evaluation systems of the OECD-DAC and its member countries follow these principles. Aid Effectiveness charters such as the Paris Declaration (2005), the Accra Agenda for Action (2008) (see OECD, 2008a) and the Busan Global Partnership (2011) have consistently emphasised the importance of *managing for results*. The UN General Assembly's review of RBM at the United Nations (2008) underlined that, notwithstanding some conceptual and practical problems, RBM still remains the dominating operational framework of the development industry.

Results-based management is linear, methodical, and provides transparency and controls in management processes (Ramalingam, 2011). The results framework is conceptually intuitive and presents a logical hierarchal chain of results, each feeding into the next level (see Figure 2).

Figure. 2. Illustration of a results chain.



Source: Kusek and Rist (2004)

Such frameworks provide the basic blocks of the *results matrix* used by most development agencies. The Canadian International Development Agency (CIDA, 2008, p.3) defines a result as a "describable or measurable change that is derived from a cause-and-effect relationship". Consequently, a result statement outlines what a policy, program, or investment is expected to achieve or contribute to (CIDA, 2008). Each result is defined by the type of change (operational, institutional, behavioural, societal), and the time-frame it takes to achieve it (short, medium, long-term). At the lower levels of the results chain the agency has more control over what happens, but as one moves up the results ladder to outcomes and impact, attribution and accountability become more fuzzy, as more factors come into play that affect the development results. Outcomes are the stage where interventions interact with the complexities of people and the real world, hence it is more difficult to establish the firm casual effect emanating from the project - this will be elaborated in more detail in the sections 2.09-2.14 on impact evaluation.

In recent years, CIDA (2009) has broken down outcome results into immediate outcomes, intermediary outcomes, and final outcomes, characterised by different time-frames and control-levels. Immediate outcomes are the easiest to assess for causality and correlation with development intervention. This has made it easier to conduct impact evaluation of agency interventions, allowing for a decent measure of change in the beneficiary population, while still allowing a fair degree of isolation and attribution.

The results framework was further enriched by the introduction of risks and assumptions, which allowed for further analysis and testing of the robustness of the program theory. Another useful framework for this type of analysis is Mark Friedman's (2005) results- based accountability (RBA), or outcome-based accountability (OBA), which focuses on well-being at the endpoint. A useful distinction that the RBA offers is *population accountability*, well-being of the whole populations (which is dependent on the operation of many stakeholders and forces) and *performance accountability*, which relates to the well-being of the customers or clients of a specific programme, agency, or service (Eisenbruch, Blignault, & Harris, 2005). With this model, specific interventions can be held directly responsible for improvements in the well-being of beneficiaries of the intervention.

While being a widely used framework in the development industry, RBM is also frequently criticised. Ramalingam (2011) argues that even in the best- case scenario, such approaches can only be followed loosely, because real-world systems cannot be divided up and controlled in neat and tidy ways. Results-based management is often criticised for being technocratic, reductionist, mechanistic, and presenting a linear representation of reality, which is normally not the case. Bhola (2000) argues that change is not only logical but also political. Results-based approaches to development management often underestimate the multiple factors, actors, and forces, both internal and external to the project, that influence the outcomes of interest. Interventions may or may not work in different contexts, and even in

the same setting projects evolve and are not static; therefore, the deterministic approach of RBM is not appropriate for complex and changing environments (White, 2013). Also Rogers (2000) explains that the logic model is not adequate for assessing, what she defines as, complex and complicated programmes, which have more elaborate and less linear theories of change.

Administratively also RBM presents many problems. As donors and supervisors often use it as a mechanism for compliance and control, development managers often undertake programming where results are readily predicted or risks are low. It encourages a formalistic approach to achieving outcomes and stifles innovation, flexibility, and experimentation. At other times RBM is just a burdensome administrative exercise where impossible results are formulated but never achieved. The UN General Assembly (2008) external review of RBM in the United Nations system revealed that RBM has been, to most UN staff, an administrative chore of little value to accountability and decision-making. Attainment or non-attainment of results had no discernible consequences on resource-allocation or human resource decision-making. Reporting on results is always very subjective as it lacks credible, impartial, and transparent methods of verification.

In spite of their flaws, logic models and results-based frameworks remain the backbone of modern public policy and development management, and thus are essential to understand and refer to when conducting evaluation of any development intervention. Revisiting the program theory and results matrix are the starting points in any evaluation exercise, and the type of evaluation being chosen is dependent on the aspects of the logic model which are being assessed. The interaction of RBM and evaluation approaches will be elaborated upon in the succeeding section.

2.4. Typologies of development evaluation

The current development arena presents a myriad of organisations each with their own goals, priorities, and agendas which can range from poverty reduction, social inclusion, protection of human rights, peace-building, assistance to the distressed, capacity-development, improving living conditions, and other causes. One challenge facing the evaluation field is that different institutions follow different paradigms of development and have different underlying objectives and principles; therefore, the way evaluations are conducted and interpreted can vary significantly (Donaldson, 2005). Even within one organisation, different projects might have different champions and managers, and might be assessed by different analysts and evaluators who operate with different frameworks and models, therefore, making it difficult to aggregate results and observe the overall contribution of even a single agency (White, 2005).

In such a setting with multiple objectives it is imperative to be clear what precisely is being evaluated and what are the questions that are being posed (Glennerster, 2009; Robson, 2007). The questions that are being asked about a programme or policy is ultimately what guides the type of evaluations being conducted. Evaluations can vary significantly depending on their purpose, the audience they are intended for, and the conceptual framework they operate in. Interventions can be evaluated under a multitude of lenses, for instance client satisfaction, coverage, achieved results, un-intended results, implementation process, quality of intervention, coherence and coordination, contribution to national priorities, and cross-cutting issues (i.e., gender, participation, capacity development). The Organisation for Economic Cooperation and Development (OECD, 2000) Evaluation Network, for example, uses five criteria for evaluating development assistance, namely relevance, effectiveness, efficiency, impact, and sustainability.

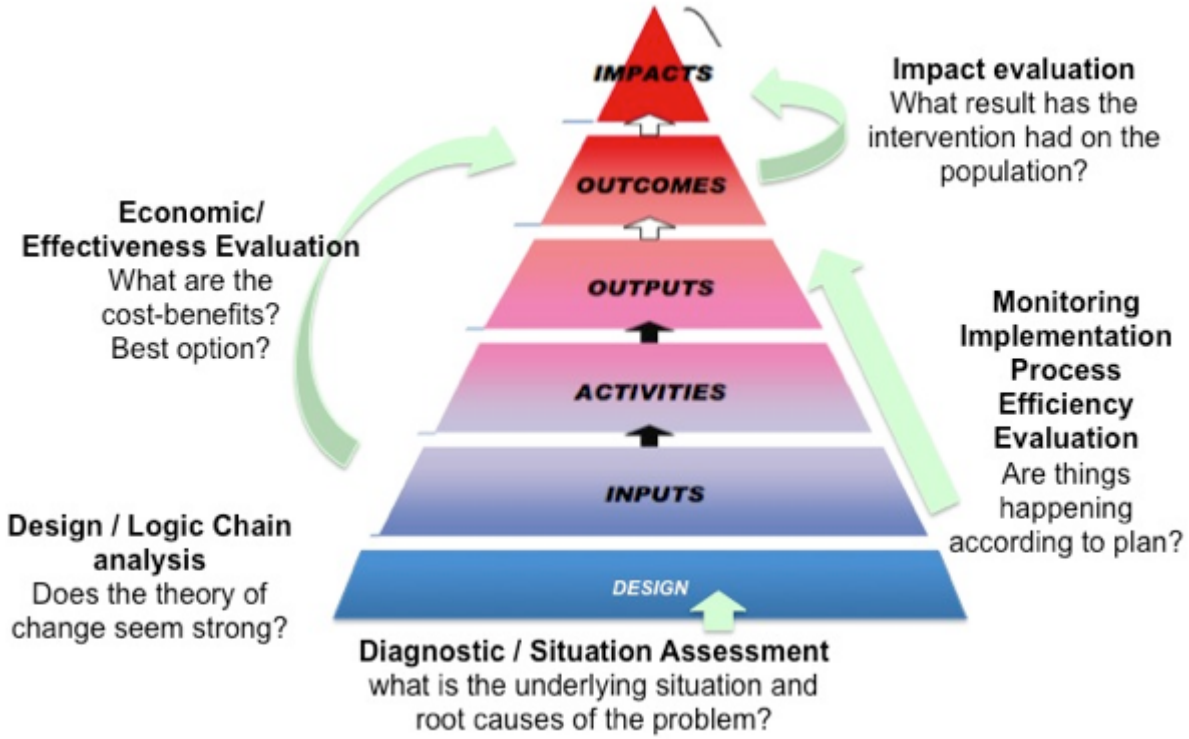
Several evaluation scholars (Kusek & Rist, 2004; Owen, 2007; Patton, 2008) have identified a number of different approaches, forms, and types of evaluation. The type of evaluation that is chosen is typically linked to the purpose it is intended to serve and the timing when it is done in the programme cycle. Much of the early evaluation theory was heavily influenced by the approaches and terminologies being used in the education assessment literature. Scriven (1991), for instance defined evaluations as either 'formative' or 'summative' in their purpose. Formative evaluation is usually done internally while the programme is still in progress, with the purpose of improvement, learning, and refining the specific development product or service. Monitoring and process evaluation is thus often associated with formative evaluation. Summative evaluation, on the other hand, is typically done at the end of the intervention to evaluate its final value, worth, and effectiveness for the benefit of external parties. In recent years, Patton (2011) has made a case for a new 'developmental evaluation' to be used for innovation, change, and learning in complex and evolving interventions typical in development settings.

These basic evaluation categories have further been sub-divided into more typologies of evaluation. Owen and Rogers (1999) suggest several evaluation forms including proactive, clarificative, interactive, monitoring, and impact. Kusek and Rist (2004) illustrate other evaluation approaches such as performance-logic chain, pre-implementation, case study, rapid appraisal, process implementation, impact, and meta-evaluation. In 2011 the South African Department for Performance Monitoring and Evaluation (DPME), published an Evaluation Policy Framework (see DPME, 2011) that included seven types of evaluations to be utilised in the South African public sector, namely diagnostic, design, implementation, impact, economic, and synthesis. Patton (2008) in his book *Utilisation focused evaluation* offers a range of evaluation designs that can be used in different contexts and for different

purposes. Depending on the questions being asked, the context, and type of analysis being undertaken, different methods and approaches are used by the evaluator in assessing the development intervention.

With respect to the results chain, discussed in the preceding section, a different form of evaluation is required depending on what point of the results hierarchy is being analysed, as illustrated in Figure 3.

Figure 3. Diagram representating the types of evaluations to be utilised at different levels of the results chain



Source: DPME (2011)

Monitoring, process and implementation evaluations usually look at the elements that are controlled by the agency such as inputs, activities, and outputs. Impact evaluation looks at the development changes (outcomes and impact) that occur in the beneficiaries as a result of the intervention. Economic analysis and CEA (Levin & McEwan, 2001) look at whether the costs (inputs) of the intervention outweigh its benefits (i.e., progress made on outputs or outcomes). Diagnostics (DPME, 2011) and situational analysis are usually done when a programme/policy starts and are meant to inform the design of the intervention. Design/logic chain analysis can be done at any moment to assess the TOC of a programme and if this is strong and still relevant.

Some of these forms of evaluation will be elaborated in more detail in subsequent sections, although the focus of this research will be on the models, methods, and approaches used in impact evaluation and comparative meta-analysis, that will guide the evaluation framework proposed in this thesis.

2.5. Validity and data quality

Before discussing the various approaches to evaluating development effectiveness, it is important to reflect for a moment on the issues of validity and data quality as this affects all the types of evaluations that will be discussed in the rest of the thesis. Different methods contain different degrees of validity and reliability, and therefore it is important to understand such implications when utilising each type of evaluation form. Identifying the potential biases and endeavouring to reduce them becomes even more important when conducting impact evaluation as will be discussed in the relevant sections.

As in all social research approaches, evaluation is caught in the quantitative-qualitative debate (Bryman, 1984; Mertens, 2014; Sale, Lohfeld, & Brazil, 2002). When frameworks are concerned with economics and issues of causality (see more in sections 2.9-2.14 on impact evaluation) the evaluations tend to be more quantitative, and when they operate within a socio-anthropological context they tend to use more qualitative methods. Evaluation is also caught into the age-long divide of objectivity versus subjectivity, and in the internal-knowledgeable or external-independent evaluator dichotomy (Christie, Ross & Klein, 2004; Conley-Tyler, 2005). The external evaluator paradigm favours objectivity, independence, and accountability and the internal evaluator paradigm is more orientated towards capacity-building, improvement, and a culture of learning (Arbab, 2000; Meinzen-Dick, 2004).

As in other forms of social research, the issue of validity applies also to the field of evaluation. Validity is what gives robustness and authority to the study (Creswell & Miller, 2000) In quantitative research validity is associated with rigour, evidence, reliability, accuracy, and objectivity and in qualitative research it is associated with trustworthiness, credibility, transferability, dependability, and confirmability (Badenhorst, 2007). Two terms frequently referred to in impact evaluation are internal validity, which is about causal inference and how much the results can be actually attributed to the intervention, and external validity which is associated with the idea of generalisability and how much the findings can apply to other context - places, people, and time (Ellis & Levy, 2009). External validity is concerned with how the results can apply to the broader population. Sampling usually plays an important role in both external and internal validity (see sections on impact evaluation).

Of great importance to any monitoring, evaluation, or social science research undertaking is the availability of quality data. When data is processed, organised, and interpreted it becomes information,

knowledge, and evidence, which in turn informs policy-making where the lives of millions of people are affected. Data is critical to measure baselines and progress being made on development indicators (see the section on monitoring). When conducting evaluations of interventions, development analysts require a range of data on development conditions of beneficiaries at the lowest geographic level (i.e., district, city, village, household, individual) and with as much time frequency as possible. There are, however, always financial, time, and logistical limitations to acquiring the ideal data required for a study, so compromises always need to be made.

Evaluators often collect secondary data through review of administrative documents of government and public entities, and official statistics gathered by statistics offices, research institutes, specialised agencies or international organisations, like the UN and the World Bank, through population census and surveys. Such document review process save a lot of time; however, such data might be incomplete or inappropriate to use because it might have been produced for other purpose and by other people who might have not applied the same rigour in the data collection or sampling. In complex evaluations analysing data from different sources can also be problematic as different organisations might use different data management systems that do not link up to one another (Rankin, 2012).

Collecting one's own primary data is preferable; however, this could have significant time and cost implications. In the production of evidence, a large range of data collection methods can be used which range from more rapid and informal, to more structured, formal, and rigorous methods that lead to more reliable, credible, and valid data (Kusek & Rist, 2004). Such methods can include interviews, observation, focus-group discussions, questionnaires, surveys, census, and field experiments (Bouffard & Little, 2004). Each of these methods have different implications for costs, time, expertise required, response rate, and level of intrusion; therefore, they need to be chosen appropriately for every given evaluation exercise.

2.6. Monitoring and the role of indicators

Monitoring is one of the most basic forms of programme evaluation. In many cases one cannot conduct other evaluations without having previously done good monitoring. Impact of a programme, for example, cannot be assessed before making sure that the programme has been implemented correctly (White, 2013). More complex evaluation approaches often rely on data that is collected throughout regular monitoring of programmes and policies; therefore, the following section will discuss the key role that indicators play for monitoring progress and later also for more elaborate forms of impact and economic evaluation.

Kusek and Khatouri (2006, p. 16) define monitoring as a "continuous process of collecting and analyzing information to compare how well a project, program, or policy is being implemented against expected results". It looks at trends and signals that things are changing in the situation as a result of the intervention. It is usually an internal management process associated with controls and performance. It acts as an early warning system, alerting managers to issues and problems, thus allowing opportunities for programme adjustments and mid-course corrections. Monitoring therefore has more of a formative role (Scriven, 1991) providing real-time feedback to managers to improve their work, adjust their course of actions, and learn as the project moves along (Bamberger et al., 2010). It encourages ownership, team-building, and group learning. It helps in regularly reviewing capacity of stakeholders, checking assumptions, and validity of the TOC at different stages of the project (Kusek & Rist, 2004).

Crucial tools used in monitoring, but also in more complex forms of impact evaluation are performance indicators, used to measure the change and the status of development results at different stages of a programme. Indicators are measurements/instruments that enable decision-makers to assess progress towards achieving an intended output, outcome, or goal (CIDA, 2008). Kusek and Khatouri (2006, p. 166) define an indicator as "a variable that provides simple and reliable evidence about the achievement of a specific result". These are particularly important when measuring outcomes, which are more intangible, external, behavioural, and institutional oriented. Indicators can cover a range of different issues and can be either pre-designed/standardised or proxy/indirect indicators (Kusek & Rist, 2004). For them to be useful they need to relate to the development framework used by the aid agencies or by the country under analysis. Many indicators have already been developed by international institutions (i.e., UN, World Bank, OECD) to measure the progress on development objectives and the global commitments such as the MDGs or the Paris Declaration. Some commonly used indicators in development economics include Gross National Product, Human Development Index, GINI coefficient, etc.

Development agencies use a mix of both qualitative and quantitative indicators. Quantitative indicators are usually considered stronger and more reliable as they allow for the use of more empirical methods of analysis. To sharpen the indicators Doran (1981) advocated for indicators to be SMART (Specific, Measurable, Attainable, Relevant, and Time-bound) and Schiavo-Campo (1999) called for CREAM indicators (Clear, Relevant, Economic, Adequate, and Monitorable). When measuring programme effectiveness, a good performance indicator should be able to respond to change caused by the intervention while isolating itself from other factors that influence the development context (Kusek & Rist, 2004). Perrin (1998), in his discussion around performance measurements, encourages the development of indicators through participatory stakeholder processes and the systematic testing and refining of such instruments as they get used.

Every indicator is usually accompanied by a baseline, which is the value of the indicator at the beginning of the intervention, and a target, which is the expected/desired value of the same indicator at the end of the intervention or any given point in time (Kusek & Rist, 2004). Indicators, baselines, and targets are the key elements required for performance monitoring. Monitoring frameworks often also include other columns such as sources of information, means and frequency of verification, and party responsible for collecting the specific data on that indicator.

Indicators are also extremely important for impact evaluations. They provide the parameters to assess if the intervention has actually made a change in the development conditions of a community, and to measure the intensity of the change that has taken place. Indicators are thus the core elements used in pre-post evaluations to assess progress towards achieving intended results as well as in comparative studies. In econometric analysis, indicators also become the variables of the statistical equation. For all these reasons, Bamberger et al. (2010) recommend evaluators to be involved from the beginning in construction of the project monitoring systems and appropriate indicators, ensuring that *a priori* the right data is collected in the right way, to address the requirements of a potential impact assessment later.

If basic monitoring is not properly done in the early stages of the programme, subsequent impact evaluation could be flawed. Before observing changes in the outcomes, managers and evaluators need to make sure that the inputs have been properly used to implement the right activities that in turn have produced the desired outputs. White (2013) explains that often impact evaluators fail to collect indicators across the results chain and to check if the intervention was actually implemented according to plan. Assumption and changes need to be checked before conducting summative impact assessments. White (2013) proposes a *funnel of attribution* model, where before impact evaluation is undertaken, the evaluator assesses basic information about the target beneficiary population: how many people actually participated in the programme, how many understood it, and how many implemented the treatment correctly? Once the evaluator has checked that the original plan was properly implemented, impact of the programme on the development outcomes can be finally assessed.

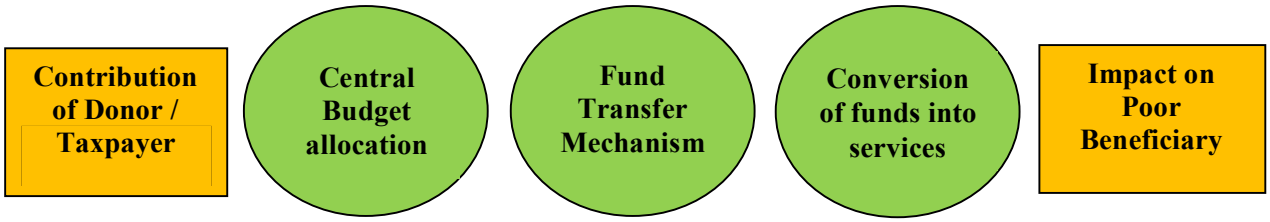
2.7. Inputs and resource tracking

For issues of accountability towards donors, taxpayers, authorities, and beneficiaries, one of the most widespread forms of evaluation is the reporting on financial inputs and resource-allocations (Trialogue, 2005; Wiesner, 2005). This is usually likened to routine accounting and auditing procedures, and therefore can be produced fairly easily and quickly. Almost every organisation in fact, out of obligations to its stakeholders, needs to account for its spending. Typical annual reports of organisations which handle development finance indicate how much money has been given, for what, to whom, and where?

This information is usually gathered from financial accounts and often analysed and presented through descriptive statistics, tables, and graphs. Some think-tanks and international organisations, like the World Bank, the UN, and the OECD, collect such information from different countries and agencies and do further macro-level analysis on aggregates and trends.

A major concern in the management of development funds is the issue of inefficiencies, corruption, and *leakages* that occur throughout the public system of both industrialised as well as developing countries. Some stakeholders and analysts ask the question: how much resources actually trickle down to the final beneficiaries? When undertaking such resource tracking analysis, it is useful to utilise an adapted version of the Filmer-Hammer-Pritchett model (see Jack & Lewis, 2009) to explain the transformational chain that occurs from the donor to the recipients of development interventions as illustrated in Figure 4.

Figure 4. Transformation of funds into results in the public and development sector



Source: Adaption of Filmer, Hammer & Pritchett model (Jack & Lewis, 2009)

This is the basic sequence of activities that occurs in an apparatus of the public, development, or humanitarian sector. This can occur at a country-level, but if it is taken to the international aid system, more layers of complexity are added as institutions from different countries and multilateral agencies are also involved. This process occurs not only within the governmental but also in the private, the non-profit, and the religious sector.

Resources invested by donors for development and charity purposes will typically go through the hands of many intermediaries before they reach the final beneficiaries. In every step of the process outlined in Figure 4 there may be different degrees of transfer costs, legitimate or illegitimate, which can be considered as losses and leakages. This concept was initially explored by economist Arthur Okun (1975) who discussed the *leaky-bucket* experiment. He described how a portion of the money from the rich to the poor will inevitably dissipate in inefficiencies related to administrative costs, tax collection, transfer systems, reduced work effort, and change in socio-economic attitudes. Okun (1975) explained that up to 15% of leakages can be expected in a public system.

The phenomenon of the leakage of development and humanitarian funds is accentuated even more so in the international arena as reported by numerous critics (i.e., Calderisi, 2006; Dichter, 2003; Easterly 2007; Maren, 2002; Moyo, 2009) who comment on the financial losses which occur in the delivery of aid, such as costly services and products, expensive delivery mechanisms, high transaction costs, organisational overheads, corruption, etc. Some prominent studies on these issues include Jepma's (1991) analysis of *tied aid* and ActionAid's (2005) report on 'phantom aid'. According to Easterly (2006), 40-60% of medicine sent to West Africa is stolen before it reaches the clinics and sold in the black market. Although there is no clear consensus on how much exactly trickles down to the final beneficiaries, commentators at the Monterrey Conference on Financing for Development (2002) (see UN, 2003) talked about only 30% of donor funds actually being utilised for their intended purpose.

More accurate and empirical studies have been undertaken by the World Bank using analytical tools such as the public expenditure tracking surveys (PETSs) and quantitative service delivery surveys (QSDSs), typically conducted in sectors such as education and health (see Dehn, Reinikka, & Svensson, 2003; Lindelow, 2003). Through the careful review of documents, official records, targeted interviews, and systematic questionnaires, PETS looks at the flow of resources through layers of public administration (i.e., the inputs and the outputs), while QSDS is more concerned with the local-level institutions (i.e., schools and clinics) which transforms funds received into services to the population. In the QSDS, issues of accountability, staff management, incentives, and quality of services are being addressed. Because of the propensity by corrupt officials to misreport, data is often triangulated and validated from multiple angles, which include users, staff, and administrators at different levels. Both PETS and QSDS have been very useful in tracking different forms of corruption, absenteeism, funding delays, inequitable allocations, misuse of services, and institutional inefficiencies (Reinikka & Svensson, 2006). Such type of country analytical exercises can be expensive, time-consuming, and require a high degree of political support, but can also lead to drastic institutional changes such as in the first PETS conducted in Uganda's education sector in 1996 (Reinikka & Svensson, 2001). One of the biggest challenges of doing such evaluations externally is the resistance of organisations to provide sensitive financial information, which is often manipulated. Quantifying non-financial inputs and services (such as in-kind and technical assistance) is also difficult to undertake when accounting organisation's development efforts (Besharati, 2013c). Assigning monetary value to the quality of products and services can also be problematic especially when conducting comparative analysis and CEA (discussed further in ensuing sections).

2.8. Limitations of development agency evaluations

Development is not only about transferring resources to the poor with the least amount of cost as possible (i.e., throwing cash from an airplane to villages in Africa) but also about the transformation and change that occurs in groups, communities, or countries as a result of an intervention, programme, or policy. The performance of an intervention or an institution needs to be therefore evaluated against results it shows on the conditions of its beneficiaries. Inputs and processes are surely important to monitor but what matters the most at the end of the day is achievement of results. This goes beyond just calculating the net outputs of a programme but measuring the progress of development outcomes in a specific population (i.e., the improvements in the living standards of poor populations), following the introduction of a development activity. The following section will explore some of the trends with regard to the way mainstream development evaluations occur and some of the shortfalls and limitations common to the evaluation of development programmes.

Historically, development institutions have focused their analytical efforts mainly on monitoring their inputs (what they spent), activities (what they did), and outputs (what they visibly produced). This has led to more efficiency, process, and implementation -type evaluations (Kindornay & Morton, 2009). Focusing on lower levels of the results chain is easier because these elements are more *controllable* by the agency (Kruse, 2003; Masud & Yontcheva, 2005; Dehn et al., 2003; Roche, 1999). Traditionally most agency evaluations, in fact, limit themselves to assessing if things have been carried out according to plan and if the organisation has performed well. Arguably, this can be partially attributed to a stronger emphasis being placed by donors on financial accountability than on actual development results (Fowler, 2013).

Although accurate information on inputs, activities, and outputs is important, such evaluations do not tell whether or not progress is being made toward solving a problem, and often the problems remain even after projects are completed (CIDA, 2012). Most evaluations conducted at the micro (project) level do not provide much information on the impact these are having at the macro (sector) level, and the effects these are having on the socio-economic conditions of the region (Pitman, Feinstein, & Ingram, 2005). Less emphasis has been placed on evaluating outcomes and impact of programs as these are always more difficult to measure. Furthermore, most development agency evaluations have been dominated by qualitative methods, which have been very context-specific, subjective, and leaning towards more positive biases (Riddell & Kruse, 1997). Programme evaluations, in fact, are often conducted by agency personnel or by external evaluators that want to maintain their good relations with their client agencies, and therefore less inclined to report negative outcomes. Also most of the respondents interviewed during agency evaluations tend to be staff, partners, and beneficiaries directly

involved in the project, and therefore likely to provide positive biases about the intervention (Bamberger, 2009).

In the late 1990s several of the Nordic agencies (NORAD, DANIDA, SIDA) commissioned prominent evaluators such as Riddell (1995), Kruse (1998), Oakley (1999) to undertake meta-evaluations of the various NGO projects around the world under their respective portfolios. Riddell and Kruse (1997) were further asked by the OECD Development Assistance Committee (DAC) Expert Group on Aid Evaluation to review evaluation reports from 240 projects from 13 case studies in donor and recipient countries, by collecting data from both reports and key interviews. The meta-evaluation of all these NGO activities found that almost all project evaluations reported that the donor objective had been fulfilled, though these conclusions were reached on fairly subjective grounds. Most evaluations, in fact, relied heavily on qualitative techniques - interviews in 90% of cases, document review on 70%, and direct observation in 60% of cases. Quantitative pre-test and post-test assessments could not be made due to the lack of baseline information in 70% of the cases. Evaluations were generally conducted rapidly and were questionable in terms of their independence from the implementing agency. In a separate analysis, Bamberger et al. (2010) concluded that most agency evaluations are conducted with very small budgets, in very short time-frames, and have not been planned properly at inception, therefore fail to collect even the most basic baseline data. As will be discussed in more detail later, for impact evaluations to be implemented counterfactuals are essential; however, in evaluations of most development programmes counterfactuals are largely absent, due also to the prohibitive costs, extensive time, and specialised technical expertise required.

There are various opinions among experts on the usefulness and validity of standard agency-level programme evaluations. Easterly (2006) has favoured micro-level evaluations over macro-level studies and encourages a continuous process of assessment of development interventions within each setting. Collier and Dollar (2002), on the other hand, criticises project-level evaluations for being costly, ineffective, and inadequate for measuring performance of organisations. Roche and Kelly (2005) also agree that very few agency performance assessments can be considerate scientifically reliable.

2.9. Measuring impact and the challenge of attribution

The development community has gradually moved beyond evaluating intervention based on their funding objective to evaluating the impact and outcomes of interventions and their contribution to the MDGs (Wolfensohn, 2005) and now the Sustainable Development Goals (SDGs). Impact evaluations are very important evaluations used in the policy space often to assess pilot programmes, and to establish their worth before they are replicated and expanded to the whole country.

Soon, however, evaluators realised that the higher the one analyses in the results hierarchy, the more difficult it is to control what happens and to attribute the development result to the intervention of a particular agency (Leeuw & Cooksy, 2005; White, 2005,). Many internal and external forces are at play for change to occur. One of the biggest challenges in development evaluation (White, 2005) is that of attribution - to be able to estimate and isolate the effects of the intervention from other factors, processes, and events affecting the same population (Rossi, Lipsey, & Freeman, 2004). Development change is usually caused not only by one programme but by the activities of many agencies and stakeholders operating in that same context. Furthermore, endogenous and exogenous forces, such as natural disasters, economic shocks, migration, epidemics, can detract from the gains achieved through the most well-intended development efforts (Johnson & Lamdany, 2005). Du Toit and Mouton (2013) argue that claiming attribution in the social sciences and in development management is practically impossible.

It is important to acknowledge that outcomes and impact take a long time to manifest themselves (sometimes even decades after the project) and different types of interventions have different maturation periods (Clemens et al., 2004). Also most interventions have a significant impact on the population during and shortly after they are implemented, due to the excitement and attention given to the recipient population during a project life-span (Schollar, 2015); however, often the developmental transformation is not sustained into future years. For this reason, it is useful to distinguish between short-term effects, occurring shortly after the intervention, medium-term effects, and long-term effects, occurring for example 5 years after the interventions (Harris, 2009). It is thus very useful to conduct impact assessments shortly after the project ends but also at later stages (i.e., 3, 5, or 10 years later) to check if the results of a development intervention have been consistent and sustained over time. Lastly, the effect of an intervention in one population can be significantly different from the effect of the same intervention on another population, hence context becomes very important. This idea will be explored in more detail in chapter 5 of this thesis.

The term *impact evaluation* has a slightly different connotation for development managers and for economists. For the traditional M&E practitioners impact evaluation is the evaluation of the last ring of the results chain - the long-term social, economic, environmental changes that a programme hopes to effect. This could include the analysis of the intended or unintended results of a development intervention. Du Toit and Mouton (2013) define this type of analysis also as *outcome evaluation*.

For economists the term impact evaluation is not necessarily correlated to the impact label of the results-matrix. As a matter of fact, most impact assessments are done on what is described in the logic model as outcomes. Impact evaluation in the economics arena is more closely linked to the notion of measuring causality and the net effects of an intervention on an outcome of interest. Once again

quantitative indicators are used as empirical measurements of the improvement in the development condition of a population affected by a particular intervention.

Economists conducting impact evaluation are concerned with the *counterfactual* (White, 2009) to answer the following question: what difference would there be in the beneficiaries with and without the programme? In practical terms, this implies the creation of a control group that is identical (or at least very similar) to the treatment group that receives the intervention. Assuming that all the other endogenous and exogenous factors remain the same throughout the process, the raw difference between the programme group and the control group indicates the impact (or the net effect) that an intervention produces. In the language of econometricians, if Y_1 is the condition of the community after a development intervention and Y_0 is the condition of the same community without the intervention, the net impact of an intervention is the difference between those two parameters [1].

$$Y_1 - Y_0 = \text{effect of the intervention} \quad (1)$$

Impact evaluation is devoted to examining whether a program, treatment, or intervention caused a particular change in the outcome of interest (Trochim & Land, 1982). Theoretically, a counterfactual would imply the creation of a perfect clone of the individuals or communities being examined. Assessing the impact of a programme is done by measuring the difference between the clone that received the treatment versus the one that did not (Gertler et al., 2011). As this is realistically impossible, impact evaluation uses complex statistical techniques to create two groups (programme group and comparison group), which are identical in their characteristics except that one receives the intervention and the other does not. Cook and Campbell (1979) explain that to conduct impact evaluation three conditions of covariation, temporal precedence, and no plausible alternative must be met before a cause-effect relation in a development intervention can be inferred.

The next sections will explore how the main evaluation approaches – econometric, experimental, and quasi-experimental – achieve these conditions when assessing the impact of development programmes and policies. For different types of impact evaluations described there might be different type of threats to validity and risks of biases. There are also different ways to reduce these threats through preventative measures, good evaluation design, rigorous methods, and statistical techniques (Trochim, 2006), which will be elaborated in the ensuing sections.

2.10. Non-experimental and econometric approaches

The non-experimental approach is possibly the weakest of all the impact evaluation methods in terms of internal validity as is vulnerable to numerous types of biases. It does not use a counterfactual; therefore, it cannot estimate precisely the impact of an intervention, because of the influence of so many other factors that can affect the development outcome (Gertler et al., 2011). Nevertheless, non-experimental methods are fairly simple, intuitive, and easy to implement, therefore even inadvertently, they are the most used methods of impact evaluation in the development industry.

The most classical methods for assessing impact of development programmes are longitudinal before and after evaluations where baseline data is collected on beneficiaries before an intervention is introduced and then the same indicators are looked at again after the programme has been completed, in order to observe any change which has occurred in the development condition of the beneficiary. At first this could seem like a straightforward and intuitive approach to assessing impact; however, the method does not take into account other factors – natural, environmental, economic, social – which might affect the outcome and therefore might inflate or deflate the impact being observed (Todd, 2012). A similar approach is taken also in Interrupted Time-Series (ITS) Analysis where panel data from multiple time-frames are looked at to observe drastic changes in a trend as a result of an intervention or event (Davies, 2012). Hartmann et al. (1980, p. 543) define ITS analysis as "a statistical method for analysing temporally ordered scores (time-series data) to determine if an experimental manipulation, an intervention, or even a serendipitous intrusion, has produced a reliable change in the scores (or trend)". Such method is, however, far from being precise and requires an exhaustive analysis of any alternative explanation for the change in the trend before the hypothesis of the impact of an intervention becomes acceptable (Davies, 2012). The ITS analysis belongs to the *reflexive comparisons* family and is often used to analyse the impact of a government policy (Baker, 2000).

The other common non-experimental method used in simple impact evaluation is 'cross-section analysis', where outcome data from the treatment group is compared with outcome data from another similar group or the general population. This type of analysis does not need a baseline as both groups are examined at the same point in time (usually at the end of an intervention). The weakness of this method is that analysts are not able to know with precision if there were already pre-existing differences between the two groups such as motivation, capacity, or any other advantages which caused the programme group to perform or underperform compared to the comparison group (Todd, 2012).

Both of these two non-experimental impact assessments can be further strengthened through the use of regression estimators (Todd, 2012), where additional data on other influencing factors can be inserted in the model to function as control variables. For this reason, the most long-standing traditional

methods to measure impact used in development economics have been various types of regression analyses. The standard OLS linear regression formulas have been variations of the following [2]:

$$y = \alpha + \beta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 \ldots + \delta_n x_n + \varepsilon \quad (2)$$

where y is the dependent variable which stands for the post-intervention development outcome, while the independent explanatory vectors are α representing the beneficiaries initial development condition, β_1

Such econometric models have been produced for both micro-level evaluations (of the impact of agency programmes on communities) but more at macro-level (for the impact of national policies and donor funding on developing country economies). Micro-level models tend to be more precise as they contain fewer factors to control for than macro-level ones.

In macro-level impact studies the most popular evaluations have been country-level or cross-country econometric models using mostly cross-section and panel data (Neumayer, 2003). Various types of multivariate regression analysis - Ordinary Least Square (OLS), Two-Stage Least Square (2SLS), Generalised Least Square (GLS), two-part model (Chenery and Strout, 1966), non-linear, and binary models - Probit, Tobit (Goldberger, 1964; Tobin, 1958), and Heckman (1979) model - have been used as instruments in these type of studies. Such research has mostly examined development interventions in relationship to economic growth of recipient countries (i.e., Guillaumont & Chauvet, 2001). Often aid-growth models have been expanded to include correlations with many other variables which affect country economic performance such as trade, macro-economic policies, financial regimes, exports, savings, governance, corruption, military spending, etc.

Many cross-country econometric models have been conducted. Rosenstein-Rodan (1961), Svensson (1999), Alesina & Dollar (2000), Brumm (2003), Easterly & Levine (2003), Roodman, (2008), have all made attempts to assess whether aid is truly making a difference in underdeveloped countries or not. Some studies have found positive relationship between official development assistance (ODA) and growth (Levy, 1988; Marris, 1970; Papanek, 1972; Riddell, 2008), but the larger portion of research concludes that aid has really little or no developmental impact on recipient countries (Boone, 1994; Doucouliagos & Paldam, 2009; Griffin & Enos, 1970; Mosley, 1980) and sometimes even has a negative impact (Knack, 2001; Moss et al., 2006; Weisskopf, 1972). Collier (2008) and Woods (2008) have found that aid is effective only at certain threshold of the country's budget or GNI. Since the late 1990s, there has been a new strand of econometric research, which has found that development assistance has impact only under certain conditions such as good governance, fiscal stability, peaceful environment, non-fungibility, trade openness, and other macro-economic settings (Burnside & Dollar, 2000; Guillaumont &

Chauvet, 2001; Neumayer, 2003). These findings, particularly the study by Burnside and Dollar (2000), have been highly influential in the development community, encouraging international institutions and bilateral donors to adopt more selectivity in their allocation of aid to countries with good policies and strong institutions. This has opened up a lot of controversies and political debates between donor countries and beneficiary countries around conditionalities (i.e. Accra High Level Forum on Aid Effectiveness, 2008) (see UNDG, 2008).

Over the years econometric models have met with growing scepticism (Deaton, 2009). The literature is still indecisive, and the methodology is not very precise and robust (Howes, Otor, & Rogers, 2011). Such methods require the availability of extensive and accurate datasets, which are not always available. Although fairly useful to assess external validity, regression estimators are very weak when it comes to internal validity, as they also suffer from sample selection bias (Dehejia, 2013). Econometric approaches will never be able to capture the full range of extraneous variables and potential factors, especially the unobserved ones, which could potentially have some confounding role on the final outcome of interest.

Typically, econometric studies review the effect of development interventions within a 2 to 5-year time-frame; however, the analysis of impact of more complex development policies often necessitate decades, like in the case of education and governance programmes (Arndt, Jones, & Tarp, 2006). By the same pretext, Clemens et al. (2004) have argued that most of the previous econometric studies have been flawed. They have provided a useful categorisation of types of development assistance and the time-frame for the results to be expected: (a) humanitarian/emergency aid that can be evaluated in the short-term as it has to do with immediate consumption (food, tents, medicine); (b) medium-term assistance, typically infrastructure, budget/balance of payment support, aid to productive sectors (agriculture, industry, and enterprises) which can be measured in the 4-year span; and (c) long-term assistance such as interventions in governance, environment, and education which requires almost a decade before impact can be properly assessed. Recent impact evaluations done in developing countries (Andrabi, Das, Khwaja & Zajonic, 2011; Evans, Kremer & Ngatia, 2014) have also shown that effects of interventions may vary depending on when they are meaured and are generally not sustained over time. The same phenomenon was presented also in the impact evaluation of the Primary Maths Research Project (PMRP) in South Africa (Schollar, 2015), which will be discussed in more detail in chapter 5.

2.11. The micro-macro paradox

In the vast econometric literature on international development, Mosley (1986) discovered an interesting paradox that was later confirmed by Boone (1996). The paradox was that when evaluating micro-level

development interventions of different agencies, the great majority report resounding success but when conducting macro-level econometric studies on the effect of all interventions in the country the results tend to be very disappointing; in other words, the aggregate impact of development programmes is less than the sum of its parts. This micro-macro paradox assumes that either (a) there are some errors in the regression models; (b) there are negative factors in the provision of aid (i.e., the aid curse) which cause negative consequences on recipient countries (i.e., inflation of economy, the Dutch disease, brain drain from government to international agencies, indirect fuelling of corruption, despotism, and conflicts); or (c) aid-receiving Governments are weakened by transaction costs of dealing with multiple development agencies (Howes et al., 2011). Some of these phenomena have further been analysed by other analysts such as Easterly (2007), Collier (2008), and Moyo (2009).

In order to test Mosley's theory, Robert Picciotto (2006), former head of the World Bank Independent Evaluation Group (IEG), conducted a study comparing the rating of the Country Assistance Evaluations (CAE) of the Bank's Country Assistance Strategies in 55 countries with the aggregate ratings of the individual project evaluations within the Bank's portfolios in the same countries. The results showed that in one-third of the cases there was a disconnect between an unsuccessful country strategy and a majority of successful sector projects, or vice-versa. This led to the conclusion that either there were some serious problems with causal chain in World Bank programming or that the micro-macro paradox did actually exist.

The Micro-Macro paradox also shows also that although the majority of the micro evaluations tend to report positive results (see section 2.8), macro-level economic evaluation demonstrates that the total effects of all the development interventions are usually negative; thus, the aggregate impact of development interventions is less than the sum of its parts. This confirms the previous assumptions that aside from the actual development intervention, many other external factors, such as socio-economic, demographic, and environmental forces, as well as multiple other development players, institutions, and macro-policies affect the development conditions of the recipient population.

2.12. Randomised control trials

Since the early 2000s there has been a boom in the use of quantitative approaches in development evaluation and a re-discovery of randomised experiments (REs) or RCTs to conduct impact evaluations of development interventions. Randomised control trials are the new frontier of impact evaluation methods promoted by organisations such as the Jameel Poverty Action Lab (J-PAL), Innovations for Poverty Action (IPA), International Initiative for Impact Evaluation (3iE), the Network of Networks for Impact Evaluation (NONIE), the National Bureau for Economic Research (NBER) and the World Bank.

In the last decade there has been an impressive rise of experimental evaluation of social programmes in developing countries, with numerous RCTs being conducted especially in Kenya, China and in India (i.e., Murnane & Ganimian, 2014; Piper et al., 2016; Evans & Popova, 2015; Bold et al., 2013; Duflo et al., 2012; Lucas et al., 2014; Glewwe & Kremmer, 2006)

Randomised experiments are often considered the most robust and reliable form of scientific evidence (Baker, 2000), the 'gold standard' among impact evaluation methods (Ellis & Levy, 2009). Randomised control trials bring rigour, high internal validity and precision to the study of development change by introducing the idea of social experiments; in other words, treating the development intervention on a particular community as a scientific experiment to be monitored closely. Randomised experiments bring the principles of controlled lab settings to social settings by using cause and effect principles and modifying independent variables to observe effects on a final outcome. Experiments are used to measure treatments (interventions) and their ability to produce an impact (effectiveness) and the extent of the change (efficacy) (Treweek & Zwarenstein, 2009).

Experimental designs, which are historically associated with the natural sciences, medicine, and psychology, are also increasingly being used to measure international development programmes, in areas of public health, education (Friedman, Kremer, Miguel, & Thornton, 2011; Hanna & Linden, 2009), criminology (Farrington & Welsh, 2005), social work, governance (Olken, 2005), and microfinance (Banerjee, Duflo, Glennerster & Kinnan, 2009; Karlan & Zinman, 2010). Some of the biggest proponents of experimental methods, so called *randomistas*, have been Michael Kremer and Alaka Holla from IPA and the various economists at the Center for Economic Policy Research (CEPR) and J-PAL, such as Esther Duflo, Rachel Glennerster, and Abhijit Banerjee.

We have seen earlier that in non-experimental approaches, impact is assessed through observing a sample of the population pre-treatment and post-treatment. The difference between those two data points deduces the change that occurred as a result of the intervention. Such a simplistic approach does not, however, take into account the many other factors which could influence change and which are not related necessarily to the treatment. A good experiment, therefore, asks the question what would have happened if the intervention did not occur? This is referred to as the counterfactual. A statistically rigorous counterfactual is developed by creating a control group that is identical to the treatment group in all its features except for not receiving the particular intervention. The classic diagram of an experiment consists of four basic elements namely (a) time (maturation of the change), represented by a linear model; (b) presence of the treatment, represented by an X; (c) observations and measures, represented by O; and (d) groups, represented by R, N, and C, depending on if a group is created through randomisation, non-equivalent group design, or cut-off point design (explained in next section 2.13). A typical two-group pretest-posttest randomised experiment will therefore be annotated (Figure 5) as follows:

Figure 5. Annotation of a simple pre-test post-test randomised experiment.

Treatment	$\boldsymbol{R \quad O_1 \quad X \quad O_2}$

Control	$\boldsymbol{R \quad O_1 \quad\quad O_2}$

Assuming that the two groups are exactly the same at start, the observed difference in the post-test between the control and treatment group can therefore be attributable only to the intervention (Kumar, 2005). Means in the post-test results can be then analysed using standard statistical techniques such as ANOVA, *t*-test, and *f*-test to check for significance. The results of the RCT can then be inserted into an OLS regression model where the impact of the intervention becomes the beta coefficient and other variables of interests can be added to the model. This basic experiment is often modified depending on the conditions and expanded along the four elements (time, treatment, observations, and groups) in order to increase validity and strength of the design. Sometimes an experiment is designed for a number of different treatments that are compared against one another and against the control group, which does not receive any treatment. This is often referred to as multi-arm experiment (McEwan, 2014).

This assumption of identical control and programme groups is however a big assumption that characterises experimental trials. How do we ensure that the treatment and control groups are actually identical so they can be properly compared? When groups in fact are created through purposeful sampling there is always a selection bias, which undermines the entire experiment. In theory a perfect counterfactual is thus created through randomisation. Randomisation theory works under the principle of *probabilistic equivalency*, balancing the biases between the groups, and creating control and treatment groups that are statistically equivalent on both observed and unobserved variables (Baker, 2000).

Experimental groups can be developed through randomisation but sometimes even through natural events beyond the control of the analyst and the participants, and therefore these are known as natural experiments (DiNardo, 2008; Dunning, 2008). When constructing social experiments, evaluators often randomise at household, group, school, or community level. This is referred to as cluster sampling (Trochim, 2006). More precise results however are achieved when randomisation is done at individual level during sampling (see chapter 5 for further discussion on cluster sampling). Randomisation implies that everyone has the same chance to either end up in the treatment or in the control group. Ideally randomisation should be done blindly (like in medical research), but in social sciences for political, practical, and ethical reasons, participants need to be informed and can choose if they wish to participate

in the treatment or not. If implemented correctly, randomised experiments can tell with fairly good precision the effects of a programme.

But when conducting randomised experiments with real people and in unpredictable social settings, many challenges and threats to the experimental design arise. These can cause major problems to the internal validity of the study (Trochim & Donnelly, 2006). One of the groups, for example, might be contaminated by receiving additional treatments from somewhere else or be affected by multiple programmes occurring at the same time on the same community. Contagion bias can thus occur when the programme and control group cannot be isolated from other interventions. Participants might decide to drop-out, refuse to participate, or be lost in the process known as attrition bias (Glewwe, 2012). Various forms of experimental biases indicate that participants behave differently if they know they are being observed. Examples of such motivational bias are illustrated in the *Hawthorne effect* (Wickstrom, & Bendix, 2000) and in the *John Henry effects* (Saretsky, 1972), where reactive behaviour, compensatory rivalry, and resentful demoralisation come into play, and participants underperform or work harder to prove themselves better than the other group. In many situations spillovers or crossovers take place where accidently or deliberately members of the control group receive the treatment or vice-versa (Glewwe, 2012). One way to counter these risks is by isolating the control and treatment groups by choosing participants who are geographically distant from each other, thus reducing the risks of spillover. However, this sometimes also makes it difficult to artificially create programme and control groups that are practically and statistically similar for the purpose of a social experiment.

Internal validity in impact evaluation is concerned with reducing potential threats, such as selection bias, design bias, interviewer bias, procedural bias, measurement bias, and response bias. Exposure bias could include attrition, contamination, spillover, and dropout. Other factors influencing the evaluation might include secular drift, interfering events, and maturation (Rossi, Lipsey, & Freeman, 2004). All of these lead to the conclusion that although considered a very rigorous method to measure impact, experimental evaluations can still suffer from dozens of different biases and threats to validity, and therefore need to be carefully controlled and performed.

There are also political and ethical problems when assigning random participants to control groups and denying them the treatment, which could be possibly life-saving or beneficial for them. To address such ethical concerns, implementers of RCTs often assign treatment through a lottery process or through an encouragement design where the programme is advertised to everyone eligible and the ones who decide to participate become the treatment group and the ones who do not become the control group (Glewwe, 2012), although this suffers also from motivational and selection bias. Other approaches used especially in the policy space are phase-in, pipeline, or wedge design, where the programme is initially

implemented with a small group to allow for experimentation, but then eventually expanded to the rest of the population once proven effective (Davies, 2012).

While being the strongest instrument when it comes to establishing internal validity (Dehejia, 2013), randomised experiments are very weak when it comes to external validity, as results are very contextual (Rothwell, 2005). The impact that an intervention might have on one group might not necessarily be the same for another group. Therefore, in public policy, the question would still remain if the results of a pilot project would be the same if the intervention would be expanded and up-scaled to the larger population?

2.13. Quasi-experimental designs

Randomised experiments might be the first-choice methods when evaluating impact of development interventions but in the real world they are extremely difficult to implement. In the social sciences and in the public policy arena there are very few situations where full RCTs can be executed. When dealing with people it is difficult to forcefully isolate program and control groups (Trochim & Donnelly, 2006) and there are many political and ethical issues that arise in such a process. Randomised experiments require advance planning, are very expensive, time-consuming, and impractical (Sanson-Fisher et al., 2007). The biggest problem however is that the overwhelming majority of evaluations in the development sector occur after the intervention has happened (retrospective / ex-post evaluations); therefore, an RCT can no longer be applied as this needs to be built-in from the start of the programme. For many interventions in the public sector the target population is already chosen, and programmes are rolled out without piloting; therefore, randomisation of beneficiaries cannot be done as this would have political consequences. Hence, randomised experiments are overall utilised more by academic researchers, rather than development managers and policy-makers.

In the social sciences and in development practice what has become more practical is the use of evaluation designs known as non-equivalent group or quasi-experimental (Campbell, Stanley, & Gage, 1963). These designs work with cause-effect relationships and have all the characteristics of true experiments (Kidder & Fine, 1987), except they are missing the key element of *random assignment*. They can be conducted quicker, cheaper and have the great advantage of being implementable after the intervention has occurred, therefore making them suitable for retrospective studies (Baker, 2000). Almost all quasi-experimental methods, however, rely heavily on the availability of large datasets collected through previous surveys or census exercises. This can make the process easier but, at the same time, riskier, as this requires locating archival records and collecting data from secondary sources, which might be fallible, incomplete, or used for different purposes. Quasi-experiments are less robust and less

internally reliable than RCTs, however much more practical, flexible, quicker and inexpensive to implement, and therefore more appropriate for a dynamic development environment.

In the absence of randomisation, quasi-experiments artificially create a counterfactual that resembles as much as possible the treatment group. To strengthen the internal validity of the impact evaluation, programme and control group need to be as similar as possible. The biggest concern in non-random designs is that of selection biases which might affect the results of the experiment (Trochim, 2006). The major frameworks for quasi-experimental evaluation designs (QEDs) can be found in the literature of Campbell and Stanley (1966), Cook and Campbell (1979), and more recently of Rossi, Lipsey, and Freeman (2004). Some of the most popular quasi-experimental methods are discussed in the following sections. Each type of design has its advantages and disadvantages, and situations in which they are better suited.

2.13.1. Matching

The most intuitive way to create control groups similar to the treatment groups is through a process of *matching*; in other words, finding participants for the comparison group which match as closely as possible participants of the programme group on key characteristics (i.e., age, education, income, gender, motivation, ethnicity, geographic area, and social class). The assumption is that each of these characteristics could in one way or other affect the outcome of interest; therefore, it is important that to the largest extent possible the treatment and control group are balanced and similar. The problem with matching is that often this is done on observable characteristics. Many unobservable variables, however, might also exist which could threaten the internal validity of the experiment and increase the risks of bias.

Some impact evaluations have utilised complex statistical matching techniques such as propensity score matching (PSM) (see Dearden, Reed, & van Reenen, 2006; Jalan & Ravallion, 2003; Peikes, Moreno, & Orzol, 2012) to overcome this problem. The idea of PSM is that rather than matching on the basis of all potential variables matching can be done on the basis of a scalar propensity score (Rosenbaum & Rubin, 1983). The method takes a number of different variables that individuals have, weighs them and collapses them into one numerical index. The propensity score is the probability of being in the treatment group given you have characteristics $X = x$ (Caliendo & Kopeinig, 2008). The approach of PSM is that for every individual in the treatment group, another person in the comparison group is identified based on a number of variables that predict participation in the treatment. Propensity score matching method relies on large samples and the use of parametric methods such as Logit and Probit to calculate propensity. Matching is usually done on individuals in the *common support area*, where the scores of the people in the control and programme group overlap. Propensity score matching utilises different techniques such as calliper matching, interval matching (Dehejia & Wahba, 2002),

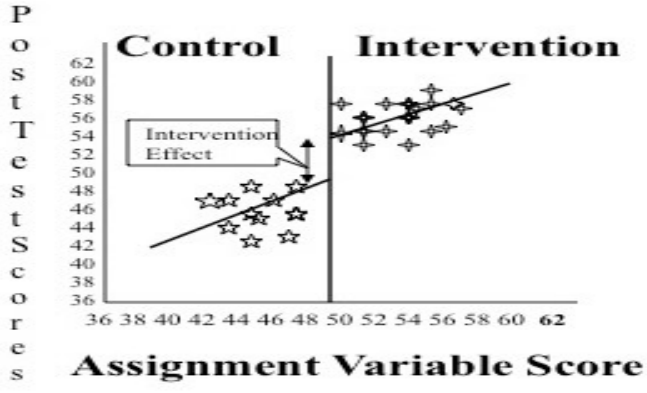
kernel density estimation (Heckman, Ichimura, & Todd, 1998), radius matching, one-to-one method, but the most popular is the nearest neighbour method – matching individuals in the control and programme group who have the closest propensity score.

2.13.2. Cut-off point design

Another popular quasi-experimental technique used in assigning participants to treatment and control groups is the regression discontinuity design (RDD) or cut-off point design (Cook & Campbell, 1986; Malamud & Pop-Eleches, 2010; Meredith & Perkoski, 2015). This technique is often used in social policies and programmes where grants, scholarships, and other targeted assistance is provided on the basis of a continuous variable such as age, test score, or poverty index. It involves choosing a clearly defined parameter/selection point (i.e., age, income, test score) and placing the individuals just above the cut-off point in say the control group and the ones just below the cut-off point (also called discontinuity) in the programme group. The theory assumes that the group of individuals who entered the programme who are just below the cut-off point are not much different from the group of individuals who are just above the cut-off point who did not enter the programme. Their participation was thus almost by chance (quasi-random), therefore producing a fairly strong and valid counterfactual. This assumes also that there was also no spillover or attrition over time between programme and control groups. Samples also need to be large enough so as to ensure that the distribution of the means of the observed covariates in the two groups are statistically similar and significant (Todd, 2012).

The impact of an intervention can be observed by a drastic change in the trend caused around the cut-off point (i.e., introduction of a programme). A graphical representation of RDD method is presented in Figure 6.

Figure 6. Graphical representation of the regression discontinuity method.



Source: Shadish et al. (2002)

Regression discontinuity design is possible when units can be ordered along a quantifiable dimension that is systematically related to the assignment of the treatment. It takes advantage of the pre-existing design of social programmes to create control groups without needing to confront ethical and political challenges. Although this technique might appear defeating and counter-intuitive at first, it actually rules out selection biases through strict group allocation formula, thus making it a strong research design. Lee and Lemieux (2010) explain that if in a RDD people cannot manipulate their eligibility scores, the treatment can be considered to be even *locally randomised*, and therefore almost experimental. For this reason RDD method is very strong in internal validity. In the absence of an RCT, one of the strongest quasi-experimental techniques used in impact evaluation is the RDD (Dehejia, 2013; Shadish & Cook, 2009). Nevertheless, as the effect is measured among individuals just around the cut-off point, results might not be generalisable to the entire population (Davies, 2012); therefore, like the RCT, it is weak in external validity.

2.13.3. Instrumental and control variables

When groups are not established through randomisation, there are serious validity threats as there are always unobservable differences between program and control groups that undermine the cause-effect links in the experiment (Campbell, 1991). In quasi-experiments there might be other plausible rival explanations, other endogenous or exogenous factors, which determine the change in the population regardless of the effects of the intervention. Other confounding variables (i.e., economic shocks, geographic location, income, education and culture) need to be therefore minimised and controlled to strengthen internal validity. Quasi-experiments often utilise a whole series of statistical controls, inserting in the equation covariates or control variables to statistically adjust the model to other potential influencing factors. This is done through techniques such as regression estimators (see section 2.10), general linear models, and the analysis of covariance (ANCOVA) (Field, 2009; Glewwe, Kremer, Moulin, & Zitzewitz, 2004; Jacoby, 2002). These techniques allow for noise reduction during experiments so as to observe more clearly the cause and effect relationship of interest. The extraneous confounding variables could be many, so it is important to choose covariates that have significant impact on the final outcome of interest.

One of the econometric methods leading to strong internal validity in quasi-experimental impact evaluation is instrumental variables (IVs) (Dehejia, 2013; Hombrados & Waddington, 2012). The idea behind IVs is that in the absence of randomisation, people self-select to be part of a treatment/programme or not. The participation into the treatment can sometimes be dependent on an exogenous variable (e.g., distance to the programme, interest in certain activities) that is correlated to the probability of

participation, but not directly correlated to the outcome of interest (Y). If an appropriate exogenous factor (IV) is identified this can be regressed in two steps with the participation variable and later with the final outcome, in order to extract the exogenous part of the decision from the self-selection so as to infer the effect of the programme due to the exogenous component. Heckman (1995), Newey and Powell (2003), and Khandker, Koolwal, and Samad (2009) have elaborated on the application of this method.

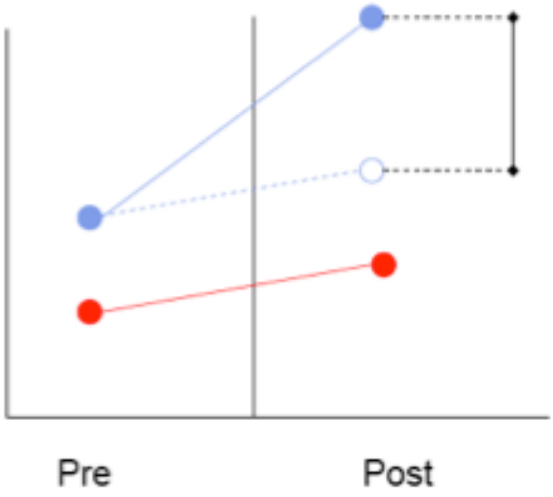
2.13.4. Double difference

Non-random designs try to artificially construct a control group based on observable variables, but there are always other unknown factors that can cause one group to perform better than the other. One way to reduce the bias and cater for the unobservable variables that exist in QEDs (especially matching) is the use of difference-in-difference (DID) or double difference (see Jacoby, 2002; Puhani, 2012; Romero & Noble, 2008). The DID method in simple terms is a combination of the before-and-after (time-series) estimator and cross-section estimator which were discussed earlier in non-experimental approaches. Double difference methods significantly reduce the threats that exist in the two simple regression estimators by combining the time variable (before and after) with the counterfactual variable (participants and non-participants). Using panel data, DID first compares data on treatment and control group before the intervention is introduced, in order to estimate the pre-existing differences between the groups, and then again the differences between the control and treatment group after the intervention (Davies, 2012). The differences of observations between the two groups in the post-test phase (endline) are subtracted from the differences of observations of the two groups in the pre-test phase (baseline); hence, the name difference-in-difference. The average treatment effect (ATE) of the intervention can thus be calculated as follows [3]:

$$\boldsymbol{ATE} = (\boldsymbol{O_{T2}} - \boldsymbol{O_{C2}}) - (\boldsymbol{O_{T1}} - \boldsymbol{O_{C1}}) \quad (3)$$

Pre-post DID impact evaluation can also be graphically illustrated (see Figure 7). By observing the difference in the post-test values of where the treatment group is and should have been, one can infer the net impact of the programme.

Figure 7. Graphical illustration of pre-post difference-in-difference impact evaluation.



Source: Adapted from Meintjes et al. (2015)

The a priori assumption of the DID method is that even though the control and programme groups may differ, they are affected by the same unknown external factors and therefore their progress and trends should be at least on the same trajectory (slope) for the fixed effects (inherent differences) between the groups to be eliminated (Glewwe, 2012). This assumption can be further tested by checking the two trends at different points of time, but this requires that extensive panel and time-series data are available. Bertrand, Duflo, and Mullainathan (2001) have also warned about the risks of the serial correlation biased in the double difference evaluation design and suggested ways to reduce such threats.

The study on the impact of South Africa's old age grant on the health of children in the households by Duflo (2003) demonstrated that the DID method could also be used with cross-section data (without time component) by comparing two sets of different variables for the same time-period (eligible and non-eligible households, younger and older girls). Differences-in-differences can also be done within estimators where differences between individuals in the same group (family, school, community) are observed over time (Rosenzweig & Wolpin, 1986; Todd, 2012), but spillovers need to still be carefully controlled.

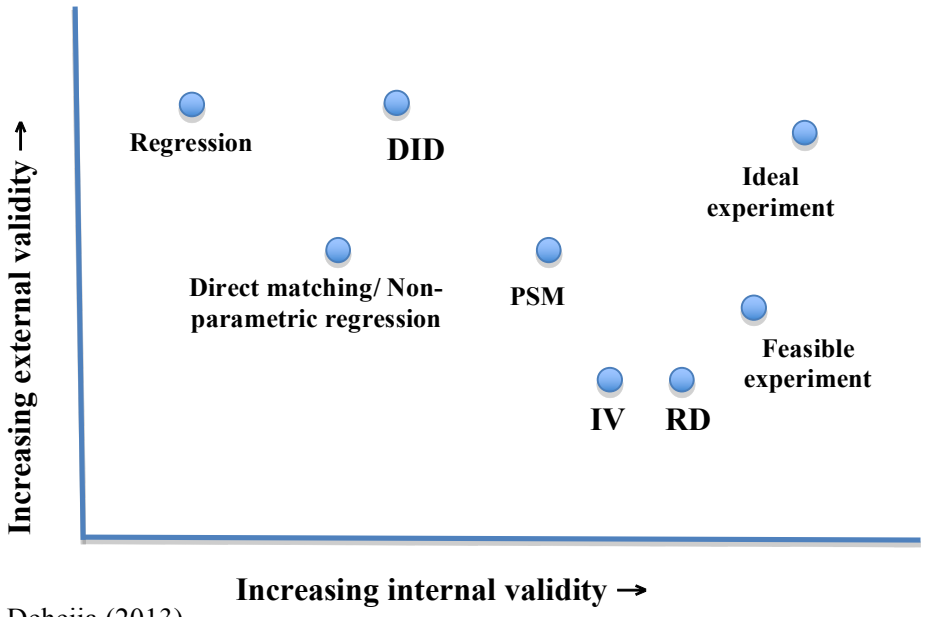
2.13.5. QEDs versus RCTs

Other non-random experimental methods include proxy pre-test design, double pre-test design, non-equivalent dependent variables design, pattern matching design, and the regression point displacement design (see more in Cook & Campbell, 1986; Trochim, 2006).

Generally, quasi-experiments are easily prone to selection biases and issues with unobservable variables, hence they require the use of complex statistical methods described above. They are not as robust and reliable as randomised experiments, but they are very flexible, practical, inexpensive, and can be implemented even many years after a development programme has been introduced. Although health sciences (medicine & psychology) still rely almost exclusively on randomised trials, economics and social sciences, and thus development studies, make extensive use of quasi-experimental methods such as PSM, double difference, regression analysis, cut-off point design, IVs, and ITS.

Although the popular mantra has been that RCTs are the 'gold standard', the truth is that different experimental and quasi-experimental methods have their strength and weaknesses, and need to be chosen carefully depending on the context, intervention, and data availability. Evidence from different meta-analytical studies (Hansen, Young, Hinami, Leung, & Williams, 2011; Lipsey & Wilson, 1993; Shadish & Cook, 2009; Smith & Glass, 1977) have shown that high-quality well-performed quasi-experimental methods, which take special care of selection bias and the construction of the counterfactual, actually yield very similar results as RCTs. Although there is no official ranking of impact evaluation designs, Dehejia (2013) has explained that RDD and IV (as close cousins to the RCT) have the strongest internal validity but weak external validity, whereas regression estimators and DID have weak internal validity but stronger external validity, while matching methods sit between the two extremes (Figure 8).

Figure 8. Matrix of experimental and quasi-experimental methods in relationship to internal and external validity



Source: Dehejia (2013)

To strengthen the validity, sometimes a combination of quasi-experimental methods is used in the same impact evaluation; for example, combining cut-off point design with regression estimators or combining PSM with DID (i.e., Blum, Krishnan, & Legovini, 2010). To validate the results of an experimental, quasi-experimental, or non-experimental evaluation, it often useful to replicate the impact evaluation using a different method, different data, and different reviewers (Dehejia, 2013), to check if the results match. McEwan (October 2016, personal communication) has also stressed that the merits of the various designs discussed above cannot be generalised, and that both the internal as well as the external validity are subject to the specific situation (i.e., data, sample size, context) in which an evaluation is undertaken. All the above discussion about the strength, weaknesses, robustness, and validity of each of the impact evaluation method will be revisited again later in section 2.18 and in chapter 5 meta-analysis.

2.14. Caveats with measuring impact

When conducting counterfactual evaluations, there will always be some degree of difference at endpoint between the indicator scores of the control and the treatment groups. It is thus important to always check if that difference is statistically significant. This can be done through a straightforward independent group *t*-test that analyses results from all the cases within the two samples. This provides a more accurate indication of whether the intervention had a real effect on the treated sample of participants compared to the counterfactual group, or if the difference was just a phenomenon of chance. While they are often used to check the results of RCTs, *t*-tests are problematic to use with quasi-experiments as the fundamental requirement of random sampling is often violated; therefore, inferential conclusions about the results cannot be validated due to the high potential of biases. Results of *t*-test, ANOVA, and other assessments of statistical significance are generally highly sensitive to sample size. It is thus critical that the largest possible samples are constructed to ensure studies are sufficiently powered so to detect the effect and allow for the most precise estimates of impact.

Many impact studies limit their analysis to whether there is a statistically significant difference between the intervention and control groups; however, this does not always indicate the direction of the impact (positive or negative) or the magnitude of the effect. Most of the basic impact evaluations, therefore, use as their *effect size,* the regression coefficient or the raw difference between the outcome scores of the treatment and control group to quantify the net effect of an intervention. This process,

however, can also be deceiving (see more in section 2.18). No matter how precise and meticulously computed, impact measures ultimately are still just estimates.

The impact of programmes is also likely to be different from one beneficiary to another; therefore, what is often reported in impact evaluations is the ATE. Different segments of the population may react differently to the treatment, and often the people who participate in development programmes (i.e., the poorest, the most disadvantaged, vulnerable) are the ones who normally gain the most out of the interventions. A more accurate result of an experiment is thus the ATE on the treated (ATT). The ATT assumes that the same effect might not necessarily apply to the broader and general population (i.e., external validity). Regression discontinuity design provides an indication of the effect of the treatment on only a small group of specific individuals close to the cut-off point and therefore indicates the ATE at Cut-Off (ATEC). Similarly, local average treatment effect (LATE) is used when conducting evaluations with IVs. Randomised control trials, where participants are selected randomly and sometimes blindly, provide an intention to treat effect (ITT) (Duflo, Dupas, & Kremer, 2008). The above, together with average treatment on untreated (ATU), quantile treatment effect (QTE), and marginal treatment effect (MTE), used often in PSM studies, are all different ways impact of interventions can be estimated. Dehejia (2013) and Glewwe (2012), have elaborated on the different kinds of *impact estimators.*

Finally, depending on what point in time an impact is assessed, the results can vary tremendously. For instance, during or shortly after a programme is completed, the impact of an intervention is more pronounced on the target population. But with the passing of time, the effects wear off, other factors come into play, people converge back to the *status quo*, and therefore there is a certain degree of *decay* in the effects of an intervention (Harris, 2009). Sometimes the opposite can also be experienced, where the natural evolution of things, other interventions and factors end up reinforcing the initial treatment. The broader population slowly conforms to the desired outcome and thus a *compounding effect* is found, where the impact gradually increases after the intervention. Regardless of a positive or negative change over time, it is important, especially in comparative analysis, to report if impact values refer to the short-term (immediately after), medium-term, or long-term (more than 5 years later) effects of a particular intervention (Clemens et al., 2004).

2.15. Unravelling the 'black box'

The evaluation approaches proposed thus far adopt a positivist orientation (Neuman, 2006) striving to be scientific and evidence-based so to promote objectivity, independence, and avoid political interference that development policy is often prone to. Importance is given to empirical measures, methodological rigour, accuracy, and internal and external validity. The concern with causality and counterfactuals, and

the rise of randomistas, has thus favoured a more prominent use of quantitative methods for development evaluation.

A purely quantitative approach, however, remains dry, superficial, and devoid of the deeper social, political, and cultural complexities at play in the development setting. Du Toit and Mouton (2013), Rogers (2000), and Bamberger et al. (2010) acknowledge that experimental designs are strong evaluation approaches for addressing causality and attribution; however, they are inflexible and not always suited for complex environments. In a real-life policy setting, there could be multiple outcomes that a development intervention might be achieving and thus it becomes difficult to isolate and disentangle one effect from the other. Context can furthermore present a series of challenges when assessing precise and externally valid measures of development impact. As discussed earlier, as well as later in the South African case studies in chapters 4 and 5 of this thesis, different beneficiaries with different characteristics, in different environments, and in different historical time-frames, might react completely differently to the exact same treatment.

Already since the 1980s there has been much criticism of experimental methods as this encourages a so-called *black-box* approach to evaluation (Bickman, 2000; Chen, 2005; Chen & Rossi, 1987; Lipsey, 1993). Traditional impact evaluations are concerned with checking if a treatment works and the effects it has on the population, but they are less concerned with how the programme works and why it works (White, 2009b). In black-box evaluation (Figure 9), half of the logic model is concealed; there is a lack of attention to the intermediary components of the programme and therefore the exact process of transformation is obscured. Figure 9 illustrates the black-box evaluation.

Figure 9. Diagram of a black-box.



Source: Mouton et al. (2014)

Du Toit and Mouton (2013) define black-box evaluation as *a-theoretical*. Evaluation is done in fact with little knowledge of the mechanisms that result in the change. This can in turn cause a very distorted understanding of the programme reality (Chen & Rossi, 1983). Through straightforward experimental and quasi-experimental methods, one can conclude that an intervention did not have an impact and

therefore recommend it to be dismissed. This approach can sometimes provide incomplete and erroneous evidence that can misguide public policy and development programming. In fact, interventions fail because of many reasons: they might be designed on poor conceptual foundations; treatment might be set at too low dosage to affect outcomes; or simply programmes have been implemented poorly or incorrectly (Chen & Rossi, 1983). Impact evaluators tend to focus too much on econometric and counterfactual analysis and often miss looking at the TOC (White, 2013) and understanding what exactly happened in that specific development context. Is the programme working? For who? Where? Why? How? It is thus critical for impact evaluation to go hand in hand with *process evaluation* and the analysis of the logic model and the casual chain - from inputs to impact, with all its underlying assumptions (White, 2009b). Bamberger et al. (2010) have also emphasised that unless impact evaluation is combined with monitoring and implementation evaluation, one cannot identify if the intervention has failed in its impact because of design or implementation failures. In the real world, in fact, projects are never implemented in the exact manner they were originally conceived (Bamberger et al., 2010).

To address these concerns many experts (Carvalho & White, 2004; Chen, 1994; Deaton, 2009; Ravallion, 2008; Rogers, 2009) have advocated for *theory-based (impact) evaluation*, as a way to combine the use of rigorous quantitative methods with program theory. The process involves getting *inside* the black box and turning it into a *glass box*, *clear box*, or *white box* (Astbury & Leeuw, 2010; Scriven, 1994). Theory-based evaluation (TBE) is about unpacking the programme so that the relationships between the inner components and the causal logic can be examined. The assumptions and TOC can be analysed and the entire hypothesis upon which the programme is built upon can be tested (du Toit & Mouton, 2013). Shadish, Cook, and Campbell (2002) explain that TBE is done by unpacking the theory with all its components and subsequently measuring to what extent the cause and effect relationships in the results chain have actually occurred. This implies a combination of process monitoring and impact evaluation. Coryn (2005) provide a detailed description of the process of conducting TBE, which involves (a) an analysis of the breakdowns in the logic chain; (b) the testing of assumptions; and (c) the analysis of expected and unintended side effects. This involves not only integrating theory in project design and project management, but also in the evaluation process, including in the questions, methods, and measurements of analysis.

Quantitative methods are not the only way to address the issue of causality in development evaluation. Du toit and Mouton (2013), in fact, encourages the use of *counterfactual thinking* even when qualitative methods are being used. An interesting qualitative approach that can be a useful alternative to the classical impact evaluation methods discussed previously, is, *causal narration* or *contribution analysis* (Kotvojs, 2006; Mayne, 2008; du Toit & Mouton, 2013). Through observation and interviewing key stakeholders, this technique seeks to construct *the story* of how an intervention caused certain social

transformations to occur. The validity of the programme theory can then be argued based on the best evidence available. Leeuw and Vaessen (2009, p. 25-26) explain that

> *"contribution analysis relies upon a chain of logical arguments that are verified through identifying and investigating alternative explanations for observed impacts. This includes being able to rule out implementation failure as an explanation of lack of results, and developing testable hypotheses and predictions to identify the conditions under which interventions contribute to specific impacts."*

The steps in the implementation of contribution analysis include (a) agreeing with all stakeholders on a reasonable logic model for the intervention; (b) analysing the activities which have been implemented; (c) verifying the validity of the logic model by checking if the outcomes did actually occur as a result of the planned set of activities; and (d) assessing other factors which could have enhanced, jeopardised, or influenced the achievement of the results.

All of these theory-based models are useful evaluation approaches that help understand when and how programmes are effective, and under what conditions (du Toit & Mouton, 2013; White, 2009b). White (2009b) proposes an approach to theory-based impact evaluation which includes (a) mapping out TOC; (b) understanding the context; (c) anticipating heterogeneity; (d) conducting counterfactual impact evaluation (through experimental and quasi-experimental methods); (e) rigorous factual analysis; and (f) use of mixed methods.

2.16. Mixed methods and participatory approaches

The quantitative and qualitative marriage is now fairly accepted among European and North American scholars, but in many developing countries (like in Africa) the divide is still very wide (Bamberger et al., 2010), often deepened by the political orientation and academic training of development managers, policy-makers, and evaluators. Traditionally, financial analysts and economists tend to favour more quantitative methods, while social scientists and anthropologists are in their comfort zones when using qualitative methods to evaluate development programmes.

In impact evaluation, quantitative methods assist in measuring magnitude and distribution of effects, external validity, and statistical significance, while qualitative methods provide depth, context, texture, and help understand processes (Bamberger et al., 2010). Qualitative methods are often criticised for being subjective, time-consuming, and limiting, but at the same time they are relatively easy to use and can shed light on results of impact assessments by assisting researchers to answer questions such as

how and *why* certain changes occur or do not occur (Mouton, 1996). Qualitative tools provide further richness in understanding the complex developmental phenomenon at hand (Neuman, 2006). The integration of personal stories also enhances the human dimension of the study; thus, interviews, observations, and focus group discussions can play a critical role in reconstructing the TOC (see sections 2.15).

In the previous section, we have seen that the use of mixed methods is central for theory-based impact evaluation (White, 2009b). Mixing approaches generally adds value to all kinds of evaluation and social research, as it draws on the complementarities offered by the different methods (Bamberger, Rugh, & Mabry, 2006). Triangulation (Denzin, 2006; Jick 1979; Sale, Lohfeld, & Brazil, 2002) is useful to address the gaps and weaknesses that the quantitative and qualitative tools inherently possess (Bryman, 2007), to cross-check data from different sources (O'Donoghue & Punch, 2003), and provide further quality assurances and superior results (Johnson & Onwuegbuzie, 2004). Applications of mixed techniques is also increasingly seen in policy-oriented evaluations that combine approaches from economics, political sciences, and anthropology (White, 2008, 2009b). Better results can be achieved not only by mixing disciplines, but also by using different analysts, different datasets, and different methods. Such triangulation helps confirm findings, strengthen validity, and add further dimensions to any research endeavour.

Bamberger et al. (2010) have suggested an *iterative process* to the use of mixed methods in development evaluation. This involves conducting qualitative work before and after the quantitative impact evaluation. A preliminary field visit is usually done in order to understand the context, the TOC, and the factors/variables that could affect the intervention and outcomes. Document review is typically the first entry-point to identify the contextual and background factors affecting a development outcome and the causal logic of an intervention. In the data collection processes, when information is missing, weak, or uncertain, qualitative methods can confirm the information gathered through documentary analysis and help fill in the gaps. Endogenous and exogenous factors discovered through qualitative data collection can later be codified, scaled and transformed into numerical variables to be used in econometric analysis. The population characteristics furthermore can provide key parameters when matched comparisons groups are created during quasi-experimental designs. A qualitative field visit can also occur subsequent to the production of initial quantitative results in order to unpack how and why the project effects came about and under what conditions (Bamberger et al., 2010).

Development theories that emphasise the importance of community-driven processes, grassroots empowerment, and participatory approaches (Cooke & Khotari, 2001; Hickey & Mohan, 2004; Selener, 1997; Torjman & Makhoul, 2012) have also affected the way development programmes are being evaluated. In his discussion around TBE, White (2009b) encourages the integration of participatory

techniques in the evaluation process. This implies a close interaction with key stakeholders to better understand programme theory, implementation, and context. Approaches can span from spending a couple of days of *development tourism* in the field (White, 2009b), to the use of the wide gamut of participatory research in action (PRA) tools (Gosling & Edwards, 2003), to even embedding an anthropologist in the project locations (White, 2009b). There are clear benefits of direct experiential fieldwork and in having the same person engaged in data collection, analysis, and validation.

A participatory approach implies a shift in the research to a more critical paradigm (Neuman, 2006) that can emphasise the political potential of evaluation. In addition to gathering data from documentary sources, the researcher can seek active inputs into the evaluation from all development actors engaged in the process – the donors, the beneficiaries, the implementing agents, and other stakeholders. Moving away from the classical positivist approach, where the researcher assumes a superior, distant, and paternalistic stance in respect to the peoples he observes, the researcher can seek active participation of the stakeholders in all stages of the enquiry process. This allows for the participants to have more ownership of the epistemological undertaking, stimulating more 'collective learning' (Arbab, 2000). Knowledge is generated endogenously with the participants through a process of participatory action research (Schafft & Greenwood, 2003; Stoecker, 1999) and reflexive science (Burawoy, 1998). This also increases participant's buy-in into the evaluation process and reduces the obstacles related to ethics and consent. The process allows for the 'democratisation of knowledge creation' (Stoecker, 1999), thus facilitating collective consultation-action-reflection for social change (Arbab, 2000).

The evaluator can use the interaction opportunity availed through informal interviews, focus group discussions in the field, or through a big stakeholder workshops, to present stakeholders the initial results of the research for their feedback. This could be done through structured discussions, where participants could have the space to provide inputs directly into the research by commenting, criticising, and confirming the initial findings. This is another form of triangulation, that helps validate the results (Bogdan & Biklen, 2007) and strengthen the reliability of the study. Such type of interactive forums could also identify critical issues in the development apparatus (interventions, institutions, community) and provide recommendations for new and more effective arrangements. Thus, academic research cross-fertilises with development practice and finds practical application in public policy. Stakeholder engagement is an essential element to ensure that the data comes to life, social learning occurs, results are applicable in the real world and translated into immediate institutional and behavioural changes. If implemented correctly with appropriate principles of applied research (Neuman, 2006), findings from evaluations could have powerful policy application and be used by stakeholders in their development decision-making.

2.17. Comparative evaluations and systematic reviews

Thus far this chapter has discussed different techniques commonly used to measure impact of programmes, policies, and development activities. This study is, however, concerned also with the process of comparison of different types of interventions and different types of institutions operating in the same sector, to assess which one is the most effective in addressing a specific development outcome. Earlier sections discussed how different development agencies follow different paradigms, frameworks, models, tools, and methods for evaluation of their programme performance (Donaldson, 2005). White (2005) has highlighted that in the evaluation literature there is not only a problem of attribution but also of aggregation of impacts of different development interventions. There is, thus, a need for a common framework that allows for meta-evaluation across various types of agencies and different types of interventions (Pitman et al., 2005).

There are some useful initiatives currently in place for the evaluation and comparison of the effectiveness of different organisations. One of them is the Multilateral Organisation Performance Assessment Network (MOPAN), which is a network of 16 donor countries united with a common interest to assess organisational effectiveness of the multilateral organisations they fund (MOPAN, 2011). Initially MOPAN conducted perception-based annual assessments, but in 2009 it strengthened its methodology by developing a common approach, which gathers over time a mix of qualitative and quantitative data through document reviews and surveys conducted with donor representatives, officials of multilaterals, and beneficiaries of the services (MOPAN, 2011). Another similar analytical exercise has been DFID's Multilateral Aid Review (MAR), which has set out to asses all the major multilateral organisations funded by the UK government in terms of their *value for money* - assessing how well they meet results while keeping costs to a minimum. The MAR methodology developed by prominent development analysts, such as Alison Evans and Lawrence Haddad, rates organisations on a composite index based on contribution to UK development objectives and organisational strengths. Although the agency ratings are quantitative, the MAR data collection and analytical framework relies on qualitative tools such as surveys, interviews, and consultative workshops with government officials and civil society of recipient countries, UK embassies/DFID offices, staff of multilaterals, and observation during country visits. DFID also conducted a similar assessment for its bilateral aid review (DFID, 2011) and the effectiveness of its various country programmes around the world.

Narrative and observational approaches, such as DFID's MAR and the MOPAN framework, provide some degree of useful comparative analysis for development policy-makers to assist in resource allocation and decisions on which interventions and institutions to favour. Nevertheless, their empirical

value can still be questionable. These type of approaches widely used in the development industry rely heavily on qualitative methods and on data collected through perceptions of diverse stakeholders, and therefore can be criticised for their subjectivity and limitations. Positivist analysts would thus argue for the use of more rigorous and empirical techniques to increase validity, precision, and scientific quality of the comparative study. In this regard, evaluation of social-economic development policy could learn a lot from the experiences and approaches of evidence-based medicine. In order to assist critical decision-making in healthcare and gather the best evidence on the effects of various types of treatments (on different populations and in different settings) a range of powerful methods are used. These include RCTs, comparative effectiveness research (AHRQ, 2012; Iglehart, 2009), cost-effectiveness analysis, meta-analysis and systematic reviews – all of which will be elaborated throughout this chapter.

Like many other evidence-driven methods, systematic reviews have been very popular in medicine, psychology, and the health sciences, widely used since the 1970s and spearheaded by the Cochrane Collaboration, a network of health practitioners dedicated to collecting and synthesising available evidence on effectiveness of treatments in clinical medicine (Gough, Oliver, & Thomas, 2012). These approaches were later adapted and expanded to social sciences through the Campbell Collaboration (CC) and the Evidence for Policy and Practice Information (EPPI) Centre. In the 1990s, the British government used systematic reviews (and rapid evidence assessments) as a tool for evidence-based policy-making (Davies, 2012), and in recent years UK-based organisations such as DFID, ODI, 3ie have been increasingly applying these methods to the field of international development (Duvendack, Hombrados, Palmer-Jones, & Waddington, 2012).

Traditional literature reviews tend to focus on only a small part of the published evidence, and are prone to a number of biases by the reviewer (selection, publication, language, subject, indexing, etc.). A systematic review, on the other hand, is more rigorous than a traditional review as it (a) follows a specific standard protocol and procedure; (b) systematically searches all published and unpublished materials; (c) is explicit and transparent on the methods and criteria used to identify, include, and appraise studies; and (d) relies on double-coding and is accountable, replicable, and updatable by other peer-reviewers. Systematic reviews attempt to identify, appraise, and synthesise all the empirical evidence that meets pre-specified eligibility criteria to answer a given research question. Researchers conducting systematic reviews use explicit methods aimed at minimising bias, in order to produce more reliable findings that can be used to inform decision-making (Cochrane Collaboration, 2014).

Figure 10. Evidence Hierarchy Pyramid



Source: www.physio-pedia.com (2016)

The purpose of systematic reviews has been to improve the available evidence, exhaustively summarise the existing knowledge of what works and how effective treatments are, and explore how interventions work on different populations and contexts (Dehejia, 2013). Systematic reviews are at the apex of the evidence hierarchy pyramid (Figure 10) as they synthesise findings across all experimental and non-experimental studies in a specific topic area. There are however instances where different systematic reviews yield diverging results, as was the case of the six reviews on education interventions in developing countries reviewed by Evans and Popova (2015) – this will be further discussed in chapter 5. This is largely dependent on the way the review is undertaken and on the inclusion criteria which determines the final sample of studies upon which the conclusions of the systematic review are based upon. The steps undertaken in a systematic review thus include (a) defining the problem and key question; (b) developing the criteria for inclusion/exclusion of studies; (c) systematic and comprehensive literature search; (d) critical appraisal of study quality and risks of biases; (e) data retrieval from the primary studies; (f) analysis, synthesis, interpretation of results; (g) presentation of findings and report writing.

The key concern in systematic reviews is making sure that 'apples are being compared to apples' (Davies, 2015). Individual studies, in fact, present usually different contexts, types of intervention, outcomes being addressed, and evaluation designs. The selection criteria for which primary studies to include or exclude in a systematic review is thus very important, and diverse literature suggest different types of categories to use. Campbell Collaboration (2015) provides a useful *PICOS* framework for

identifying the parameters based on which to include or exclude studies in a systematic review. These parameters are defined by the choices made by the reviewer with regard to a) *P*opulation or participants (this could include context and time); b) *I*ntervention (including dosage, components/features and degree of implementation); c) *C*omparator or control group (which should be similar across studies); d) *O*utcome (including measurement construct such as indicator and unit of analysis); e) *S*tudy design (including methodologies, sample size and data used).

Systematic reviews can synthesise evidence through narrative (qualitative) means, 'vote counting' (Davies, 2015), or through quantitative means (meta-analysis). While narrative and qualitative systematic reviews are able to include the largest amount of studies and explore in more depth the mechanism of change (Mouton et al., 2014), they are also subjective and suceptible to the beliefs and biases of the reviewers (Evans & Popova, 2015). A more scientific way of tallying and systematically weighing the results of different experimental studies is through the use of meta-analytical approaches, which will be discussed in the next section. Statistical meta-analysis is a key component of the methodological framework used in this research and in the systematic review of South Africa's schooling interventions undertaken in chapter 5.

2.18. Comparing effectiveness using meta-analysis

2.18.1. Traditional meta-analysis

When there is a need to empirically compare the effectiveness of different development interventions, by analysing the results of different experimental evaluations, a very useful tool is meta-analysis (MA). MA, which is an integral part of many systematic reviews, is a statistical method that simultaneously analyses, combines, and contrasts results of different studies (Hedges & Olkin, 2014; Higgins & Green, 2008). It is somewhat of an *analysis of analyses*, where large volumes of data are synthesised and observed for patterns and relationships between different studies (Borenstein et al., 2009). Individual studies all carry certain assumptions and conditions; therefore, by pooling the results of various studies together, biases are reduced and validity is increased (particularly external validity). The idea is that a MA of different studies can provide a more accurate estimation of the effect of a treatment than a single study would.

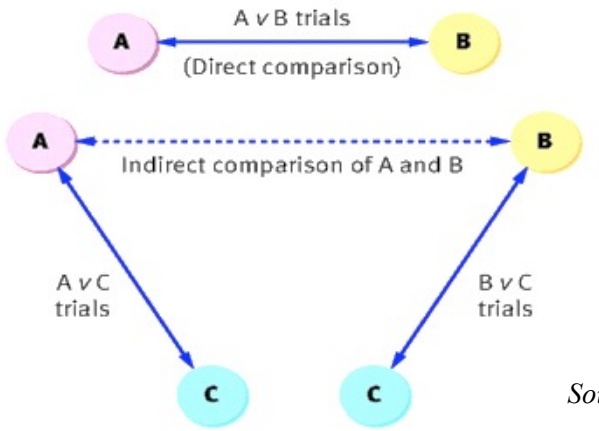
Meta-analysis has its roots in seventeenth century astronomy; however, the first identifiable meta-analytical study can be traced to Pearson (1904) in his investigation of typhoid inoculations. It was, however, psychologist and statistician Gene Glass (1976) who coined the term and contributed to its systematic use along with others such as Schmidt, Hunter and Jackson (1982).

2.18.2. Comparative meta-analysis

Specific types of meta-analytical approaches such as multiple treatment meta-analysis (MTMA), mixed treatment comparison (MTC) and network meta-analysis (NMA) are better suited for comparing effectiveness of different types of interventions (see chapter 5 of this research), All of these are extension of traditional MA elaborated by academics such as Caldwell (2005), Lu and Ades (2006), Salanti (2012) and Zoccai (2014). As opposed to standard MA, MTMA assesses and compares different types of interventions aimed at achieving the same outcome. This method helps to answer policy questions and assists decision-makers when there are diverse options of investments to make or interventions to choose from.

For example, a doctor could potentially conduct a MTMA of Aspirin, Tylenol, Paracetamol, Analgin, Ibuprofen, in order to better advise a patient on which drug is most effective to use for fighting headaches. Ideally to have a robust and reliable comparison a very large randomised experiment should be implemented, creating a control group and a programme group for every type of treatment being compared. Although this would be the best option most of the times, it is very difficult to find such types of studies, and even if they exist they would suffer from low external validity because of the restrictions in the population groups. In the absence of a large multi-treatment RCT, comparisons are made through observational means, by assessing one intervention in relationship to another intervention or a null (placebo) situation. Multiple Treatment Comparison (MTC) is often a simple head-to-head pair-wise comparison (such as comparing aspirin and paracetamol). Head-to-head comparisons, however, are often difficult to find; therefore, often meta-analytical studies are used Indirect Treatment Comparisons (ITCs). ITCs allow for comparison of interventions that have never been compared before by using a common comparator. Figure 10 illustrates the difference between direct and indirect comparisons (Song et al., 2011).

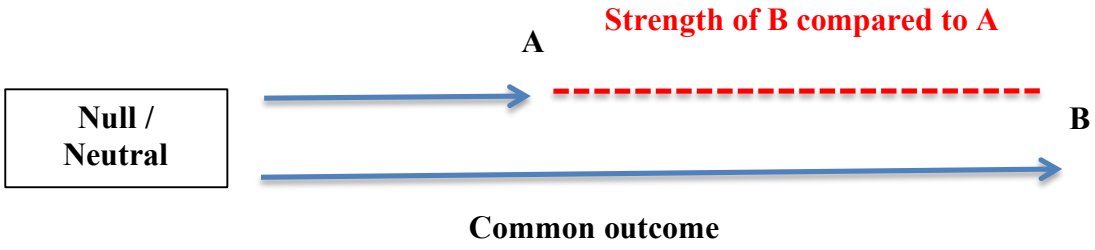
Figure 11. Diagram representing direct and indirect comparison.



Source: Song et al. (2011)

In order to do an ITC, there needs to therefore be a common comparator, whether that is active or passive (null or placebo). If using a common measurement system for the outcome (see section 2.6 on indicators) and a good non-active comparator which can function as a control group for two or more treatments, it is easy to also measure the exact magnitude of the strength of one intervention compared to another, as illustrated in Figure 11.

Figure 12. Illustration of an indirect treatment comparison



Source: Author's own compilation (2013)

This model, however, is based on the big assumption that the comparator (the control group) used to assess the effectiveness of the various interventions is similar and consistent across the various studies. In different studies, in fact, different treatments are tested on different populations that can be more responsive than others (Cranney et al., 2002). For this reason ITC, which is essentially an observational approach, is susceptible to diverse biases and therefore cannot be taken as a precise measure. Evidence has in fact shown that indirect and direct comparisons do not always yield the same results (Donegan, Williamson, Gamble, & Tudur-Smith, 2010; Zoccai, 2014) and analysts are encouraged to undertake systematic tests of consistency (Sturtz & Bender, 2012).

A more advanced form of MTMA is NMA, which effectively combines MTC and ITC methods, pooling and analysing together results from both direct and indirect comparisons. This generates a complex web of relationships between different studies, hence the name NMA. The advantages of NMA is that it not only combines different evidence but also accounts for correlation, ranks different treatments, and evaluates consistency by checking confidence intervals and *p*-values (see more in Donegan et al., 2010; Bucher, Guyatt, Griffith, & Walter, 1997; Salanti, Giovane, Chaimani, & Caldwell, 2014). Ranking of interventions, which is usually done using Bayesian probabilistic frameworks, such as

Markov Chain and Monte Carlo Cycle, provides decision-makers a flexible and intuitive tool to assess the best treatment to address a particular outcome of interest.

Whether a traditional MA or comparative meta-analysis is undertaken, the key steps undertaken in all systematic reviews, such as formulating the right question, defining clear inclusions criteria, systematic and comprehensive search of the literature, double-coding, all needs to be followed. Some key concerns, however, specific to conducting a meta-analysis are the appraisal of methodological quality of the study, the standardisation of effect sizes, and the examination of heterogeneity.

2.18.3. Study quality and risk of bias

Meta-analytical studies need to be comparable conceptually, methodologically, and statistically (Duvendack et al., 2012). The most complex exercise in the screening process is judging the methodological soundness of each individual study, deciding whether they should be included or not in the MA. The tension lies between not leaving out any useful and important evidence and making sure that a few bad studies do not influence negatively on the overall findings of the MA (Hombrados & Waddington, 2012). In deciding which studies are methodologically sound, assessments need to be made on their internal and external validity, construct measures, statistical errors, and checking if the experiment has been executed correctly.

While the inclusion of qualitative studies would be clearly very problematic, even in quantitative studies there are more than 50 potential threats to validity (Davies, 2012) which include selection bias, performance bias (spillover, crossover, contamination), attrition, detection bias, reporting bias, cofounding (implementation) bias, and motivational bias (such as Hawthorne and John Henry effect). For this reason MAs in the medical field restrict itself to the rigorously implemented RCTs and strong studies that provide high-quality evidence on the subject (Sutton et al., 2008). Higgins and Green (2008) do not recommend including studies which are not RCTs in MA. Glass, McGraw and Smith (1981), however, have argued that excluding certain studies defeats the whole purpose of systematic reviews and MA, which is to analyse all available evidence in the field, including inconclusive and unpopular studies, and what is often known as *grey literature*.

Assessing methodological quality becomes even more problematic in the social sciences and in international development where very few RCTs exist, as the field is dominated by quasi-experimental studies (Higgins & Green, 2012), which are known for their high risks of bias, and validity threats caused by the unobservable variables which might exist between control and treatment groups. Deciding which quasi-experimental studies to include or not in the MA is a complicated debate (Hombrados & Waddington, 2012). Many tools exist to assess methodological quality such as Cochrane's GRADE system and Risk of Bias (RoB) Framework and the tools offered by AHRQ, CEBP, EPOC, NICE, SIGN

50, Wells, DIAD, and Maryland. However, most of such frameworks are not appropriate for the social sciences (Deeks et al., 2003) because they are mainly designed to check validity of RCTs and epidemiological designs. Furthermore, rating systems for studies are generally very controversial, subjective, and discouraged as a source of reliable quality assessment (Juni et al., 2001; Wilson,Tanner-Smith, Lipsey, Steinka-Fry, & Morrison, 2011).

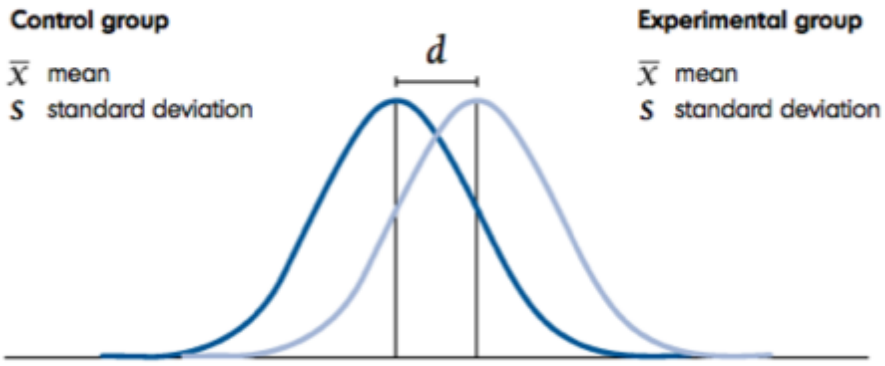
As discussed in section 2.13, high-quality quasi-experimental studies, carefully controlled with appropriate statistical techniques, yield similar results as RCTs (Hansen et al., 2011; Lipsey & Wilson, 1993; Shadish & Cook, 2009; Smith & Glass, 1977). Quasi-experimental designs (QEDs) where allocation rules are clear (Hansen et al., 2011), where selection bias have been appropriately addressed, and where the counterfactual and causality is strong, can potentially be included in MA together with experimental studies. Hombrados and Waddington (2012) suggest also that studies with exogenous selection (such as natural experiments, IVs, RDDs), and studies where time and intra-group unobservables are carefully taken care of (such as combining DID and PSM) could potentially be included in MAs of development evaluations.

Including QEDs in MA ultimately remains a judgement call based on expertise of the reviewers in both statistics and the content matter. If QED studies are finally included in a MA, it is important to assess carefully and report on the risk of bias, assumptions, statistical problems of each of the studies, so that these can be used as predictor variables in subsequent moderator, sensitivity, and sub-group analysis, as will be explained later in this section (Lipsey & Wilson, 2001, Cornell & Mulrow, 1999).

2.18.4. Standardising effect sizes

The measure of impact that is used to analyse results across studies is what is referred to in MA as *effect size (*Ellis*, 2010;* Hedges & Olkin, 2014*)*. Effect size provides analysts and policy-makers with a common translation for the direction and magnitude of the impact of an intervention (Hombrados & Waddington, 2012). It is the net impact an intervention has on the outcome of interest. It indicates how many standard deviations separate the treatment and control group (TEA, 2014), and in some of the literature it is also referred to as z scores (Rosenthal, 1994). In econometrics, it is the treatment variable or the beta coefficient of the regression. In matching-based studies it is the ATT between groups after matching and in DID it is the change of the variable of interest between groups over time (Hombrados & Waddington, 2012). The difference between the means and standard deviations of two groups gives an indication of the effect size (*d*), as illustrated graphically in Figure 12.

Figure 13. Graphical representation of effect size *d*.



Source: Authors own compilation (2014)

The challenge when pooling together results from different studies, is that different impact evaluations use different outcome indicators, measurement systems, scales, and assessment instruments, which make the exercise of comparison very challenging. When conducting a MA between impact studies it is therefore important to *standardise* results across studies to the same scale and same effect size to allow for synthesis and comparison.

There are many different measures that can be used for analysing effect size (Ellis, 2010; Schmidt & Hunter, 2014); however, it is important that within one MA all results are converted in a consistent and common manner. The *d index* is among the most popular effect size used for studies with continuous variables (i.e., test scores, income, life-years, etc.). The widely used *Cohen's d*, also known as the standard mean difference (SMD), is calculated by dividing the mean difference between programme and control group by the pooled standard deviation [4] as follows:

$$\boldsymbol{SMD} = \frac{\boldsymbol{X_t - X_c}}{\boldsymbol{S_p}} \quad (4)$$

Another effect size often used is the response ratio (RR), which is the percentage change in the outcome as a result of the programme, adjusted by the pooled standard deviation. Response ratio is intuitive and can be done with minimum data requirements. When it comes to dichotomous and categorical variables, risk ratios and odds ratios are also frequently used (Schmidt & Hunter, 2014). Odds ratio (also referred to as ratio of ratios) is the odds of success in treatment against the odds of success in the control. Similarly, risk ratio is the probability of success in the treatment relative to probability of success in comparison, and therefore it is expressed in percentage terms. Other measures used to calculate effect size include *p*-values, proportion ratio, Pearson's *r* (correlation coefficient of a regression) and U_3, which is the

probability that a randomly selected member of the treatment group will outperform a randomly selected member of the control group. Among the *d*-index measures *t*-statistic is also used; however, this can be noisy (Hedges & Olkin, 2014; Becker & Wu, 2007). Regression (beta) coefficient on the treatment dummy variable is also used; however, this can be problematic as the same covariates need to be used across all the studies in the MA.

Each of the above types of impact measures can be transformed and converted among one another, allowing for standardisation of effect sizes and thus allowing for easy comparison between studies. The decision on what effect size to use will often depend on the data availability, on the nature of the variables (dichotomous or continuous), on what is easier and faster to do, and on ensuring that the least amount of studies are lost in the process (Hedges & Olkin, 2014; Borenstein et al., 2009).

2.18.5. Data extraction and confidence intervals

No matter what type of measure is chosen, every effect size is still just an estimate, that comes with different degrees of precision, variance, and standard error. Every effect size expresses the ATE; therefore, this is always accompanied by a 'confidence interval' which provides the 'upper limit' and the 'lower limit' of the impact estimate, at 90%, 95%, or 99% significance. Standard error and confidence interval are very much dependent on the sample sizes of treatment and control group used in the study. The larger the experimental samples, the smaller the standard error and shorter the range of the confidence interval, and thus the effect size estimate will be more precise. The way sampling is done, whether random, purposeful, or clustered will also have a big effect on the standard error (Bloom, Richburg-Hayes, & Black, 2007; McEwan, 2015; Taylor, 2013). Hedges (1981) devised a formula to correct Cohen's *d* estimates when sample sizes are small, and thus this effect size is known as Hedge's *g (*Schmidt & Hunter, 2014).

In order to compute effect size, data need to be extracted from each of the individual studies. Such parameters normally include sample size (always needed to correct for bias and compute confidence intervals), mean, and standard deviation for both control and programme group. Sometimes results of *t*-test, standard error, and *r* coefficient, are sufficient to calculate effect size. Some other statistics which are often used to compute impact measures in MA include ANOVA, table of counts, *F* test, *z*-value, and *p*-value. Unfortunately, not all primary studies report all the information required; therefore, reviewers are often forced to contact the original authors of the studies to retrieve the additional statistical data. The methods and approaches of MA and the formulas to compute various types of effect sizes have been detailed by authors such as Borenstein, Hedges, Higgins, and Rothstein (2009), and Lipsey and Wilson

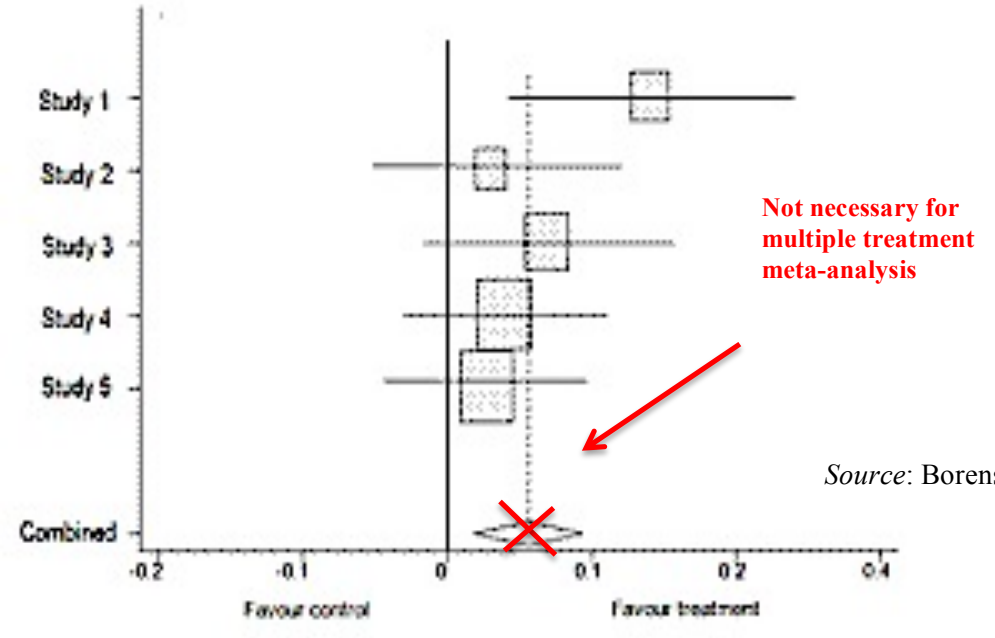
(2001). Softwares, such as RevMan, Meta XL, CMA, Stata, R, SPSS, Wilson's ES calculator, are also available to assist in these statistical calculations.

2.18.6. Analysing results of MA

A range of models are used when analysing results of meta-analytical studies (Hedges & Olkin, 2014). Fixed effect model is used when findings from different studies present similar conclusions; therefore, the aim is to achieve more precision in the estimates of effect size. In the fixed effect model, results from studies using larger samples are given more weight than studies with small samples because they contain larger margins of error. Random effect model (also called inverse variance) is used when studies present large heterogeneity of results; therefore, the aim is to estimate the actual average effect size of an intervention. Other models include the Mantel-Haenszel approach, the Peto method, and other approaches as suggested by the Cochrane Collaboration (see Chandler, Clarke, McKenzie, Boutron, & Welch, 2013). In NMA and MTMA, used also in this research, heterogeneity is obviously expected; therefore, random effects model that excludes the calculation of average pooled effect size is the most appropriate way to go.

The results of a MA are traditionally illustrated through forest plots (Figure 13) which provide readers with user-friendly graphical comparison of different effect sizes and confidence interval for each study in the MA. This also allows for an easy eye-ball assessment of the best intervention to address a particular outcome of interest, thus providing a very powerful tool for decision-makers and policy-makers (Sutton et al., 2000).

Figure 14. Forest plots (without pooled effect size).



Source: Borenstein et al. (2009)

2.18.7. Exploring heterogeneity through moderators

The final stage in most meta-analytical endeavours is the exploration of heterogeneity (Hedges & Olkin, 2014; Higgins & Green, 2008). Heterogeneity is tested in order to analyse the degree of dissimilarities between the various studies included in the MA (Hedges & Olkin, 2014; Wilson & Lipsey, 2001). In traditional MA, large heterogeneity can be a problem; however, in multiple treatment comparison (see for instance in chapter 5), heterogeneity is to a large extent expected, considering that the interventions being analysed are very different and implemented in different contexts. The analysis of heterogeneity can, however, reveal why and how certain interventions have more effect than others. By holding the outcome and the treatments constant, the test of heterogeneity will analyse if the difference between studies is due to other factors, beyond just sampling error (Hedges & Olkin, 2014; Borenstein et al., 2009). The different characteristics of the study coded during the earlier screening processes become the predictor variables in the statistical analysis of variations. These factors are often called moderators and can be drawn from the PICOS selection criteria (see section 2.17) and other study characteristics such as methodological design, population context, intervention features (Wilson & Lipsey, 2001). These moderators are usually collected and coded in early stages of the review together with the other statistical data retrieved from each of the studies.

The various moderators are statistically analysed in order to see if they have a strong influence on the effect sizes (as well as the standard errors), aside from the treatment itself. When variables are categorical, moderator analysis is usually done through ANOVA, but when they are continuous, multivariate regressions are appropriate tools that can be used (Hedges & Olkin, 2014; Raudenbush, 1994; Van Houwelingen, Arends, & Stijnen, 2002) – this process is also known as meta-regression. Other common methods used to test heterogeneity include *z*-test, Cochrane's Q (Walker, Hernandez, & Kattan, 2008), chi-square, and i-squared statistic (Borenstein et al., 2009), which provides the percentage of variability across studies not due to sampling error. Sensitivity analysis (Hedges & Olkin, 2014; Walker et al., 2008) is also performed, to check if a specific study is skewing the results of the MA in a specific direction. As a result, analyses are often repeated with and without outlier and doubtful studies, to see if results vary. When MAs present a strong and obvious degree of heterogeneity, studies are often clustered/stratified and *sub-group analysis* is conducted using again some of the PICOS sub-categories. Once the reasons for heterogeneity have been statistically explored, these can be explained in a narrative discussion with regard to caveats and limitations of the systematic review.

2.19. Cost-effectiveness analysis

This chapter has discussed extensively the concept and measurements of impact and effectiveness; however, when investors and policy-makers are making decisions around development programmes to initiate, the element of costs becomes very critical. Resources are usually finite; therefore, both donors and beneficiaries are concerned with accountability of public funds, making sure those are spent in the best way to achieve the most impactful development outcomes. This brings one back to the original question of the enquiry which is how to assess which interventions provide the biggest 'bang for the buck' or the best value for money? In the business world, CEA is done in the interest of the company's profits. In the development sector, the costs and the benefits are assessed from the perspective of the public at large, and in particular the most poor, needy, and vulnerable beneficiaries.

Impact evaluation and cost analysis of interventions are usually conducted separately using different methods and approaches; however, in the comparative framework of this research, the integration of these two processes into a CEA model becomes a very powerful tool for development policy-making. CEA combines cost-analysis results with results of impact assessments in order to ascertain and compare the effectiveness of different interventions (Glennerster, 2009) when producing change in a specific development outcome and within specific resource constraints.

Cost-effectiveness analysis belongs to the family of methods and approaches of cost analysis, which also include cost-benefit, cost-utility, and cost-feasibility (Levin, 1975). Such economic analysis tools are frequently used to assist in making decisions about appropriate courses of actions to undertake (Belli, Anderson, Barnum, Dixon, & Tan, 1998). Often different competing strategies are available to policy-makers; therefore, cost-comparison analysis is used to explore alternative options. In situations where funding is scarce, CEA can guide the process of resource allocation. There might be multiple solutions to address poverty problems; however, donors and development managers are constantly searching for the most powerful intervention that can achieve the maximum results at the lowest cost (Belli et al., 1998). Cost-effectiveness is not only used when making choices, but also when endeavouring to improve public systems, such as using limited resources more efficiently, expanding what can be achieved within a certain budget, and reducing costs to achieve a certain objective, so that funds can be invested instead in other important activities (Levin & McEwan, 2001).

The National Institute for Health and Clinical Excellence defines CEA as an economic study designed to measure effects and consequences of different interventions to achieve the same outcome measured in 'natural' units' - life-years gained, death avoided or diseases cured (Laxminarayan, Chow, & Shahid-Salles, 2006; Phillips, 2009). Like many of the methods discussed previously, CEA originated in health policy when economic analysis was used to assess what would be the best treatment to pursue to

fight a particular disease. Gradually it started to be used extensively also in the field of education (Levin & McEwan, 2001; Summers, 1994) to measure, for example, the best ways to raise mathematical skills, student learning abilities (i.e., Tan, Lane, & Coustere, 1999) or school attendance. Other applications have been in the economic evaluation of microfinance projects (Schreiner, 2003) and transport infrastructure (Gwilliam, 1997).

Cost-effectiveness analysis operates on similar principles as cost-benefit analysis, except that it is more often used when the benefits of an intervention cannot be converted into financial terms (i.e., healthy lives, higher test scores, reduced pollution). Nonetheless, these benefits still need to be quantifiable (i.e., percentage gain in development outcomes) and measured with quantitative indicators. A cost-effectiveness ratio (CER)[5] often used to compare different interventions (programmes and policies) can be calculated as follows:

$$Intervention\ CER = \frac{Cost}{Impact} (5)$$

This provides policy-makers an easily interpretable numerical value that indicates how much a programme costs for every unit of development gain. Impact evaluators, for example, often refer to the standard deviations change that an intervention causes with say $100 investment (Kremer, Brannen, & Glennerster, 2013; Taylor, 2013). Some interventions might have big impact but also be very expensive to implement, while others may have modest impact but are relatively cheaper.

As discussed in the previous section on MA, the CER would include the appropriate effect size chosen to indicate the common impact measure across different studies being compared. Experimental and quasi-experimental studies provide different impact values such as ATE, ATT, ITT, etc. Econometric studies utilise beta coefficient of the regression as the impact value (Tan, Lane, & Lassibille, 1999). When conducting comparative analysis standardisation is crucial; therefore, the use of effect sizes, such as the Cohen's *d*, Hedge's *g*, or SMD, are among the easiest and safest options of impact measures to be integrated into the cost-effectiveness model.

To further enhance precision and provide a fairer cost-effectiveness comparison of different development interventions it is also important to integrate the numbers of beneficiaries (children, households, farmers, etc.) that benefitted from the intervention into the CER equation (Belli et al., 1998). The number of participants, in fact, influences both the cost of the programme as well as the final effect expected. An intervention for a small group of people is likely to produce a bigger effect than an intervention for a large segment of the population. Size of the programme has a definite influence on the results of a CER. In an economy of scale, costs should be lower with larger programmes, but also effects tend to be smaller and more diffused among a bigger population. It is also important to look at the

duration of the programme and to see how many people did partake of the benefits of the intervention over the years. A more appropriate and reliable cost-effectiveness estimate of a development intervention could thus be calculated as follows [6]:

$$Intervention\ CER\ = \frac{(Total\ Cost)/\ (\#\ beneficiaries)}{Effect\ Size} \quad (6)$$

One challenge of CEA is that, just like in MA, CERs can be calculated for only one objective (development outcome) at a time. In social programmes besides the benefits stemming from the primary purpose of the intervention, there are often also 'externalities' (Haveman & Wolfe, 1995), which are other indirect and unintended benefits that an intervention contributes to in the community development process. A public awareness programmes, for instance, might be designed to increase prospective employment opportunities, but inadvertently might also contribute to other social benefits such as reduction of crime, family cohesion, peace-building, or maternal/child mortality (Summers, 1994). There are times, therefore, that multiple outcomes (benefits) of an intervention need to be factored in. In such cases a cost-benefit model is more appropriate, as it is used when evaluating a project or investment for all its merits against its overall costs (Shively & Galopin, 2013).

The challenge of cost-benefit analysis, however, is that all benefits needs to be standardised and monetised, and therefore be attributed a specific value. When dealing with development outcomes, valuing for instance academic achievement versus nutritional gain becomes a more subjective, politically sensitive, and less empirical exercise. Evaluators and policy-makers might feel more comfortable to conduct separate cost-effectiveness assessment for every outcome of interest rather than trying to merge everything in one composite cost-benefit model.

Having previously discussed at length the complexities around the impact measures, equal attention also needs to be given to gathering the cost data to use in the cost-effectiveness model. This can be an equally complex exercise as it requires retrieving budgetary and expenditure data, which are often inaccurate, incomplete, and difficult to retrieve from public and private institutions. Assuming that there is political will by the organisation and financial information is available through open budgets or rigorous tracking surveys, such as the PETS and QSDS (Dehn et al., 2003), the cost information is often not reported in a consistent manner.

Taking the total organisational budgets at face value, for the cost calculation of the CER model, can also be deceiving and lead to erroneous conclusions. It is thus important to disaggregate programme budgets, carefully analyse all the costs of an intervention, using the *ingredients method* (Chambers & Parrish, 1994a, 1994b; Dhaliwal, Radhakrishnan, Tsang, & Yang, 2012; Levin & McEwan, 2001), in order to ensure no important piece of financial information is left out of the calculation. Although there is

no official comprehensive list or right and wrong way to calculate ingredients, evaluators conducting CEA often include the following in programme costs: personnel, transport, facilities, equipment, materials (supplies), transfer costs, oversight/administrative costs, and other inputs required for the intervention to be undertaken. In addition to the costs to the implementer, there are also costs to the user and other stakeholders partaking, engaging, or supporting the development initiative, which need to be kept into consideration. Costs are extremely susceptible to context and may differ in different times and places. When comparing interventions undertaken in different countries and throughout different years, the analyst needs to factor in currency inflations and discount rates.

Dhaliwal, Duflo, Glennerster and Tulloch (2011), Harris (2009), and Levin and McEwan (2001) have elaborated on these considerations about CEA. The most important factor, when using CEA in comparing different types of interventions, is to be consistent in the methodology and in the assumptions being held throughout the analytical process (Baker, 2000). If performed well, CEA provides donors, managers, and policy-makers in development an additional instrument to provide useful evidence for decision-making about programmes and policies.

2.20. Conclusion: developing a conceptual and methodological framework

Some of the latest and most prominent frameworks, approaches and methods currently utilised in the evaluation of development endeavours, sourced from academia, public policy and the development industry literature have been reviewed in this chapter. Having analysed the utilisation, strengths, weaknesses and limitations of these methods, key elements from these diverse approaches can be drawn out to propose a potential conceptual framework that will guide this evaluation research and the methodological design which will be appropriate for the upcoming case studies around the effectiveness of interventions in South Africa's education sector, discussed in chapter 4 and 5.

2.20.1. Purpose and requirements

Like all evaluation endeavours, the main focus of this research is the generation of empirical evidence and useful knowledge for policy-making. The framework, therefore, needs to be practical and easily applied to development policy, assisting managers in decision-making, resource-allocation and selection of interventions to replicate and scale-up. It should enable policy-makers to choose the best interventions that will achieve the best development results while providing good value for money or return on (public) investments. Notwithstanding the above characteristics, it needs to take into account the limitations of context, time, resources, ethics that characterise the complex development landscape. Considering the

public policy implications, it is important that the framework is empirical, scientific and objective. The instruments and methods need to effectively address the issue of causality and attribution when conducting impact evaluations and strike a fair balance between internal and external validity. The evaluation framework needs to be flexible, adaptable and tailored to different geographic contexts and development sectors, so that it can be used to measure and compare effectiveness across a large variety of different interventions, agencies and institutions.

Based on the above objectives and requirements, this research proposes an evaluation framework that is appropriate and effective for the type of questions and policy analysis conducted in the development sector. Drawinging on the insights and experiences gathered through the literature (discussed in this chapter), the framework suggested for this research combines elements of CEA, MTMA, mixed methods and theory-based impact evaluation (White, 2009b). In particular, the use of quasi-experimental methods to measure effectiveness is being favoured; and a participatory applied research orientation is encouraged. The framework also integrates the use of systematic review as a precursory step to statistical meta-analysis used during the comparative analysis. The framework draws on many approaches and methods originating in the health sciences and in economics, which favour a positivist approach to social enquiry. Its foundation is the generation of empirical evidence that can inform development policy that affects the lives of millions of people. The above characteristics will form part of the criteria used at the end of the research process (see Chapter 6) to appraise the framework and judge its value and utility against other evaluative approaches prevalent in the international development system.

2.20.2. Elements of the proposed evaluation framework

Unpacking the theory of change

The framework will avoid a 'black-box' approach, often seen in impact evaluation literature (see section 2.15). From the inception and throughout the process, the analyst will endeavour to understand and unpack the intervention and its workings, and the factors and context that affect the development outcomes. The intervention theory (Chen, 1990; Levison-Johnson et al., 2009; Weiss, 1998) will be clearly established and assessed on its design strength and logic. The preliminary phase of the evaluation will assist in defining the causal chain and the TOC related to the intervention and the development problem being addressed. The programme theory (Funnell & Rogers, 2011) will be developed not only by reviewing previous literature on the subject but also through qualitative instruments (observation, interviews, focus groups with relevant stakeholders) and through techniques such as contribution analysis (Kotvojs, 2006; Mayne, 2008). Reviewing the programme implementation process through basic

monitoring will assist in ensuring that the intervention was executed correctly according to the plan before more complex impact evaluation is conducted.

The initial qualitative analysis will also assist in revealing the socio-economic information, the environmental forces, the institutional contexts and the endogenous and exogenous factors, that affect the development outcome of interest. If these factors are properly identified, they can later serve as confounding variables and moderators for the statistical controls during the impact assessment and MA. Mixed methods will be utilised throughout the research and, where possible, experiential field visits will be conducted by the analyst to understand better the context, programme theory and encourage interaction with beneficiaries and stakeholders. Positivist, interpretive and critical research paradigms will therefore merge in a theory-based evaluation approach (Du Toit & Mouton, 2013) that calls for the mapping out of the TOC, understanding the context, the anticipation of heterogeneity, counterfactual impact evaluation, rigorous factual analysis and use of diverse mix of methods (White, 2009b).

Aligned to the most popular management frameworks utilised in public policy and in development, the evaluation framework will operate against the classical backdrop and principles of result-based management (Kusek & Rist, 2004). A key step therefore will be to identify the 'outcome of interest', which the intervention is trying to address, and the evaluation is trying to assess. Useful for this analysis is the CIDA (2012) specification of immediate and intermediary outcomes, which observe the changes in social-economic conditions of communities and beneficiaries as a result of a development intervention with an easier degree of attribution. As explained by Clemens et al. (2004) and Harris (2009), appropriate time-frame expected for the maturation of results after an intervention is important to keep in mind during the evaluation process, considering that impact looks different in the short (immediately after the intervention), medium and longer term (5 years later).

Once a clear development outcome is selected careful attention needs to be given to the selection of appropriate performance indicators to measure the progress made on the specific outcome. Considering the prominent use of quantitative methods in the evaluation framework, such as MA, experiments and CEA, outcome indicators need to be quantifiable and measurable from a statistical point of view. Data quality and data availability is thus critical. Information about the intervention, about the development outcome and its indicators as well as other social-economic variables need to be available at micro-level and with regular frequency in order to utilise some of the quantitative methods discussed above.

Quasi experiments for impact evaluation

At the heart of the evaluation framework is the empirical assessment of the impact produced by the development interventions on a specific target population. This is possibly the most complex exercise as

issues of causality, attribution (White, 2005) and isolating the net effects of a programme from all the other external factors, need to be addressed (du Toit & Mouton, 2013). The framework will assist in the construction of a reliable counterfactual that can indicate the condition of the beneficiaries with and without the programme. For this type of exercise, it is important to draw from the experiences and approaches of seasoned experts and agencies in the field of impact evaluation (i.e., J-PAL, World Bank, IPA, 3ie). Although it is generally acknowledged that RCTs are the 'gold standard' of impact evaluation methods (Baker, 2000; Duflo, Glennerster, & Banerjee, 2009), in practice they come with a lot of problems related to costs, time and ethics. The biggest problem for development policy-makers however is that most development evaluations occur only after the intervention has taken place (retrospective evaluations).

An RCT can be utilised if the policy opportunity arises where an experiment can be built-in from the start of the programme. However, as the vast majority of the evaluations in the development sector are done ex-post, quasi-experimental designs will be more frequently utilised within the framework of this research. The exact type of impact evaluation method applied will depend on the type of intervention, the way it was implemented and the data available. In the absence of randomisation, counterfactuals will be created through statistical techniques used to reduce potential bias and ensure a good balance of internal and external validity. Some of the quasi-experimental methods which will be used as part of the evaluation model will include RDD, PSM, IVs, and natural experiments - each to be chosen depending on the specific context of programme implementation and evaluation setting (see section 2.13). To reduce biases and increase validity, counterfactual techniques will be combined with the double difference method (DID) to strengthen the quality of the quasi-experimental design by triangulating time and cross-group variations. To the extent possible experimental and quasi-experimental evaluations will be done using the largest possible sample sizes in order to provide enough statistical power and increase the internal validity of the results. Based on the experimental model shown in Table1, the main DID equations used in the impact evaluation framework is the following (7):

Table 1. Classical Experimental Model

$$N_T \quad = \quad N_C$$

$$N_T \quad O_{T1} \quad X \quad O_{T2}$$

$$N_C \quad O_{C1} \qquad O_{C2}$$

$$Y_x = (O_{T2} - O_{C2}) - (O_{T1} - O_{C1}) \text{ (7)}$$

where N_T and N_C are the statistically similar samples for the control and treatment population created through counterfactual methods, such as matching, RDD, IV, and RCT. In the DID model, Y_x is the impact of an intervention (x) which is the result of the difference in the observations of the treatment and control sample in the post-test (O_{T2} and O_{C2}), after they have been subtracted from the difference between the treatment and the control sample in the pre-test (O_{T1} and O_{C1}).

Econometric approaches and regression estimators (Glewwe et al., 2004; Jacoby, 2002) will also be used to control for external factors and cofounding variables (endogenous and exogenous factors) that affect the development outcome. Although less precise than the experimental and quasi-experimental techniques, econometric models will be used to provide further information and enrich the analysis of the intervention, context and development change occurring. The multivariate regression that will be utilised as part of the framework package will be different variations of the following equation [8]:

$$Y_b = Sit_b + ES_0x_0 + Cov_1b_1 + Cov_2b_2 \dots Cov_nb_n + \varepsilon, \text{ (8)}$$

where Y_x is the development outcome of the beneficiaries after the programme is implemented, Sit_b is the initial situation/conditions of the beneficiaries before the programme (baseline or pre-test), ES_0 is the effect size of the intervention (x) and Cov_1, Cov_2 ... Cov_n represent the other factors (environmental conditions, economic situation, family background, external forces, beneficiary motivation, etc.) which might have an influence on the impact model, and ε is the unexplained error term.

Integration of cost-effectiveness analysis

In development policy, cost implications become very critical decision-making factors; therefore, the evaluation framework proposes the integration of a cost-effectiveness model to assist in gathering evidence on the best option to undertake in terms of value-for-money. Once the appropriate quantitative values have been gathered for both impact and cost of programmes, these can be inserted into a straightforward CER equation that indicates the cost involved in achieving a unit change of impact through a particular intervention. For a more precise calculation, the number of beneficiaries receiving the treatment will be integrated, as this affects both costs and impact of the programme. Both of these parameters may vary greatly depending on the scale of the programme. The adapted CER [9] used in the evaluation framework can thus be expressed as follows:

$$CER = \frac{Total\ Cost\ /\ No.\ of\ beneficiaries}{Effect\ Size} \quad (9)$$

Development programmes that cost less, reach more participants, and have a bigger effect will obviously be favoured against those that are more expensive, less impactful and smaller in scope. Such indexes will allow policy-makers to compare interventions and identify which policies and programmes should be replicated and capitalised in the future.

When constructing the cost-effectiveness model, financial information will be extracted from budgets and audited financial reports of the institutions involved (donors, government and implementing partners); however, cost to the users (beneficiaries) and other partners need to also be taken into account. Costs of programmes will have to be carefully disaggregated and analysed using the *ingredients methods* (Chambers & Parrish, 1994a, 1994b; Dhaliwal et al., 2012; Harris, 2009; Levin & McEwan, 2001) to ensure that all elements have been carefully considered. Where information is not available in official reports, the reviewer will need to consult with the original evaluators and managers of the programmes, and a fair degree of estimations on missing data might still need to be done.

Standardising and comparing impact results

The numerator value of the CER is expressed in quantitative terms and represents the ATE, computed from the various types of impact evaluations discussed above. In econometric studies, usually the regression coefficient for the intervention is used, and in quasi-experimental studies using DID method, the impact value is the raw difference in the post-test outcome indicator of the treatment and control group. The challenge, however, when comparing results from different impact evaluations is that different studies utilise different instruments, scales and systems of measurement. To allow for comparison, standardisation is thus crucial; therefore, the use of effect sizes, such Cohen's d and SMD, derived from the meta-analytical literature, are some of the best options for expressing impact value of continuous variables that can subsequently be used in the comparative meta-analysis and in the cost-effectiveness model. The standard equation used to compute SMD [10] is illustrated as follows:

$$\boldsymbol{SMD} = \frac{\boldsymbol{X_t - X_c}}{\boldsymbol{Std_{p\ (t+c)}}}, (10)$$

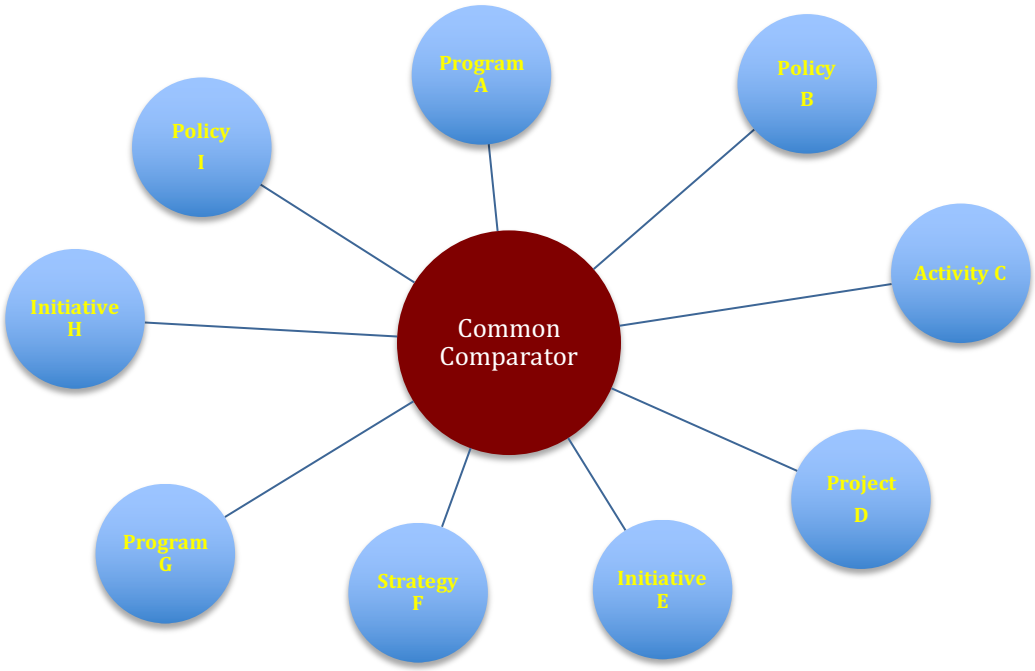
where $\boldsymbol{X_t}$ is the mean score of the treated group, $\boldsymbol{X_c}$ is the mean score of the control group, and $\boldsymbol{Std_{p\ (t+c)}}$ is the pooled standard deviation of the scores of all participants of the experiment. As discussed earlier many impact evaluations integrate DID method; therefore, statistics are provided for pre and post for both control and treated groups. In such instances the formula expands as follows [11]:

$$\boldsymbol{SMD} = \frac{(\boldsymbol{X_{post} - X_{pre}})_t - (\boldsymbol{X_{post} - X_{pre}})_c}{\boldsymbol{Std_{p\,(t+c)(post+pre)}}} \quad (11)$$

Whether large or small, effect sizes may or may not be statistically significant. Effect sizes are still in fact just estimates; therefore, they are always accompanied by other statistics such as standard error, variances and confidence intervals (usually set at 95%). Sample size has a large influence in estimating the precision of the effect size. Therefore, the larger the sample size the smaller the standard error and shorter the range of the confidence interval. Hedges (1981) devised a formula to correct Cohen's d estimates when sample sizes are small, therefore we will also use 'Hedge g' as the main effect size to compare studies and interventions (such as in the MA in chapter 5).

As the evaluation framework endeavours to provide a structured system to empirically compare different effects across interventions, the meta-analytical approaches which are better suited for this exercise are Multi-Treatment Comparison (MTC), MTMA and NMA), discussed at length by Caldwell (2005), Lu and Ades (2006), and Salanti (2012). Multiple treatment comparison tries to emulate a large multi-arm experiment, by utilising a common control group. Comparisons are made through observational means, by assessing one intervention in relationship to another intervention or a null (placebo) situation. This model is based on the assumption that the comparator used to assess the effectiveness of the various interventions is somewhat similar and consistent. In a particular country context, where different policies and programmes are being assessed with regard to effectiveness in addressing a same development outcome (such as in the case of the South African MA in chapter 5), the use of NMA can be graphically illustrated (Figure 14).

Figure 15. Graphical representation of network meta-analysis.



Source: Author's own compilation

As opposed to traditional MA that pools results from studies of the same treatment in a synthesis to improve available evidence, MTMA is used to compare different treatments aimed at achieving the same outcome. While plotting effect size and confidence intervals for each intervention is useful for observing differences between the various treatments, the combined average effects can be ignored as this is anyway based on very heterogeneous interventions. A forest plot that would thus emerge out of a comparative meta-analysis was illustrated previously in Figure 14 on page 77. To assist policy-makers and development managers make decisions and choices through simple observational comparison, forest plots could be produced first for the simple effect sizes, to observe the impact of each of the intervention, and later other forest plots could be produced to observe the cost-effectiveness of each of the interventions, using CER estimates.

As heterogeneity of results is in great part expected in a multi-treatment meta-analysis, the framework encourages the exploration of heterogeneity through *moderator analysis* (i.e., meta-regression, sub-group analysis, and sensitivity analysis), so to understand better which contextual, intervention, and methodological factors are closely linked to results of the impact and CEA, and why? Thus, coding and extraction of relevant variables from the individual studies during the early process of the systematic review becomes important so to allow the undertaking of subsequent statistical analysis of the moderators during the MA.

When conducting MA, the key concern is making sure that apples and apples are compared, and that interventions and evaluations are similar conceptually, methodologically, and statistically (Duvendack et al., 2012). The inclusion criteria for the review, therefore, needs to be clearly delineated. When comparing, synthesising, and pooling results of different studies together it is also critical to assess the methodological quality of the various primary evaluations to ensure that few bad studies do not skew incorrectly the overall results of the MA. To assist in these processes, the approaches, protocols, and tools offered by the Cochrane and Campbell Collaboration, such as the PICOS framework, the PRISM quality assessment tool, the RoB tool, and Wilson's effect size calculator, Cohen's effect size guide, will be adapted and used as part of the evaluation framework of this research (see more in chapter 5).

Mixed methods and participatory orientation

Many of the quantitative methods explained above, such as the use of impact evaluation and meta-analytical techniques, and the large quantities of data management required by this type of evaluation approach, also require a high intensity of statistical analysis, both descriptive and inferential. Combining intervention, developmental, and social-economic information creates large datasets that need to be organised, cleaned, and managed. As part of the techniques outlined above, summary statistics need to be calculated, such as means, standard deviations, and sample sizes of both treatment and control groups. Other more complex operations that need to be undertaken include significance testing, calculation of standard error, ANOVA, multivariate regression, PSM, and calculation of effect size and confidence intervals (Borenstein et al., 2009; Lipsey and Wilson, 2001). To present results in a user-friendly manner to policy-makers and public readers, tables, plots, and diagrams need to be created to illustrate results of the statistical work. Many different software packages would be required to support this complex statistical analysis: Microsoft Excel and PowerPoint, SPSS, R, RevMan, Wilson's ES calculator and comprehensive meta-analysis (CMA), which will all be used in this research.

As can be deduced thus far, the overall evaluation approach leans heavily on quantitative methods. At the same time, however, the framework is grounded in theory and relies also on mixed methods to provide depth to the analysis and strengthen both external and internal validity. The

evaluation framework will integrate elements of qualitative methods that will strengthen the data collection, analysis, and verification process, thus allowing for triangulation and superior results (Bryman, 2007). It will involve a mixture of desk review and anthropological research in the field.

Participatory approaches (Schafft & Greenwood, 2003) and applied research (Neuman, 2006) will also be utilised to engage stakeholders in the research process and to facilitate application of learning and finding to development practice, management, and public policy. Once the initial statistical analysis has been processed, preliminary findings can be synthesised, simplified and presented to diverse development partners (government, private sector, CSOs, foreign donors, academics, etc.). This will be done through small focus group discussions, presentations, or large policy workshops, where stakeholders (beneficiaries, donors, implementers, experts) will have a chance to engage with the research. Programme participants can provide direct inputs into the study, by commenting, criticising, and validating the results as they emerge. This will bring further qualitative depth to the data and allow quantitative results to be triangulated (Bamberger et al., 2010). Such process of participatory applied research will increase research transparency and stakeholder buy-in to the evaluation process, encouraging more effective dissemination of findings, and immediate use of evidence for policy change and improvement of systems and programmes.

2.20.3. An organic process of development, piloting and refinement

The development effectiveness evaluation framework being developed during the course of this research will be an ongoing work-in-progress. The model has been conceptualised drawing on the experience from the literature, using the best available theory and tools, and has benefited from the inputs of different experts engaged in the field of development evaluation, locally and internationally. The proposed framework will thus have to be field tested in different settings, populations, and sectors, in order to see if and how it works under different conditions. Insights, experiences, and learning gathered from the application of the framework can then be used to constantly improve and refine it into a more powerful and practical tool that can be used again in other development policy context.

What has emerged in the analysis is that for the evaluation framework to be effective certain preconditions are required from the development setting (policy sector and country context) that is being analysed. Firstly, the area of exploration needs to be politically significant and of high interest to policy-makers and investors, who need to provide resources and have the political will to allow for the evaluative process to occur and for data to be provided. For a meaningful comparative analysis to occur, a large number of diverse interventions and institutions need to be implementing activities in the chosen terrain and addressing the same development outcome. Outcome of interest needs to be clearly specified

and impact of interventions needs to be (at least in theory) easily attributable and observed in the short-term. Considering the type of impact evaluation and meta-analytical methods being proposed in the framework, it is critical to have quantifiable indicators for the outcome of interest, with reliable data available regularly at the lowest geographic level. To allow for the use of statistical methods and to justify the time and costs of these complex evaluation procedures, programmes and policies need to be quite sizeable and affecting a large number of beneficiaries.

Having assessed the required preconditions of use, the evaluation framework, approaches and methodologies discussed throughout this chapter will be piloted through two local studies (see chapters 4 and 5) that assess and compare the effectiveness of different interventions to improve learning outcomes in South African public schools. The research will identify which programmes and policies proved to be most successful and worth reconsidering for future programming by public and private institutions. The application of the conceptual and methodological framework in the two empirical chapters will be briefly summarised again in sections 4.4 and 5.4. The findings of the individual case studies will be discussed at the end of chapters 4 and 5. The thesis will conclude in chapter 6 with a reflection on the application of the evaluation framework to the particular context of South Africa's education sector, and what lessons can be extracted for future applications to other countries and development sectors.

Chapter 3: Contextual overview :

South Africa's education sector and interventions to improve learning outcomes

3.1 Introduction

In the previous chapters, this research has set out to develop an evaluation framework for the empirical measurement and comparison of impact of different interventions aimed at achieving the same development outcome. In order to test whether the conceptual and methodological framework is effective, practical and appropriate, it will eventually have to be piloted within a development context, common and relevant to Africa and other developing countries, to prove its worth and usefulness for real-life public policy and programme management. As discussed in the introductory chapter, education plays a central role in South Africa's social and economic development, and therefore the sector was identified as a good testing ground to pilot the set of methods and approaches for the evaluation of development effectiveness, discussed in this research. The South African education sector is also an interesting international case study, where the private sector has played a special role in contributing to development efforts, complementing the efforts of government and other traditional donors (Besharati, 2015).

Chapter two offered an in-depth review of some of the main evaluation models, approaches, methods, techniques, to measure and compare the effectiveness of development interventions. Chapter four and five (subsequent to this chapter) will illustrate the application of the evaluation framework, and the methodological package presented in this research, to two education case studies conducted in South Africa. The first is an impact evaluation of the education programme of Anglo Platinum mining corporation, stemming from its community investments in North West and Limpopo provinces. The second is a comparative meta-analysis of programmes and policies implemented in the past 15 years by different institutions to improve learner results in South African public schools.

This chapter thus provides the contextual background, institutional and policy environment, and a brief situational analysis of South Africa's education sector. An overview of the politcal, social and economic challenges inherited by the country post-1994, and the manifold investments made not only by public institutions but also by the private sector to address the education challenges of the country, will

be presented. It will explain how learner achievement has been measured in South Africa with both local and international instruments, and the poor learning results that South Africa still registers in comparison to other developing and middle-income countries (Bloch, 2009; Spaull, 2013). It will review the existing South African literature to identify some of the key societal and schooling factors that affect learning outcomes. It will try to summarise the theory of change that underpins the transformations which South Africa's education sector is currently undergoing, and which guide the manifold development interventions of various governmental, non-governmental and private educational agencies. Section 3.7.3 will delve deeper into the CSI in South Africa's education sector, which has been one of the focus areas of this research, and which will be further elaborated in the case studies in chapters 4 and 5.

This chapter does not claim to be a complete review of the complex dynamics of South Africa's schooling sector, which have already been treated in much detail by many South African education scholars (Chisholm, 2004; Fiske & Ladd, 2004; Fleisch, 2008; Bloch, 2009; Taylor, Van der Berg, & Mabogoane, 2013; Motala, Morrow, & Sayed, 2014; Sayedet al., 2013; Spaull, 2013). Much of the academic literature on South Africa's education sector focuses on the first post-democracy decade (1994-2004) of education policy development, while some of the more recent data and literature stems from the publications of various South African governmental and non-governmental agencies active in this policy space (Motala & Pampallis, 2005), such as the Department of Basic Education, JET Educational Services, Centre for Development Enterprise (CDE), Centre for Education Policy Development (CEPD), and the Human Sciences Research Council (HSRC).

As it is not the main focus of this research, this chapter will not go into a substantive analysis of the South African pedagogical literature. Many important and controversial debates that have animated South African education policy, such as access and equity, outcome-based curriculum, school fees, language of instruction, teachers and unions, school accountability, will be dealt with only marginally. The review in this chapter is more restricted and serves simply the purpose of providing a brief descriptive overview and a contextual backdrop to the subsequent empirical work to be conducted in chapter four and five, where the specific impact evaluation methods and meta-analytical techniques from the research framework will be illustrated. The conceptual and methodological approach presented in this thesis will be applied in the evaluation of South Africa's education policies and programmes, to give an example of how scientific evidence can generate useful knowledge and answer important questions that affect public policy and private investments for development, such as in the case of South Africa's education sector.

3.2 Historical and legislative context

Education has been at the heart of South Africa's political, social and economic development processes (Fiske & Ladd, 2004). The apartheid legacy created parallel systems of education where white children would receive high-quality Western training while black students would receive minimal training to function in the country's system as "hewers of wood and drawers of water" (Sayed et al., 2013), remaining subservient to the white minority (Mandela, 1994). One of Apartheid's architects, Minister of Native Affairs, Hendrik Verwoerd stated:

"There is no place for the Bantu [the native Africans] in the European community above the level of certain forms of labour ... What is the use of teaching the Bantu child mathematics when it cannot use it in practice?" (Clark & Worger, 2003, p. 48)

One of the most lasting and racially detrimental laws introduced by the Apartheid government was the 1953 Bantu Education Act (Fiske & Ladd, 2004), which put Church-sponsored African schools under State control. This would ensure that the curriculum, textbooks and teachers would enforce Apartheid ideology and undermine African culture and aspirations (Fiske & Ladd, 2004). Many independent missionary schools who until then had a played a strong role in training African leaders, the likes of Nelson Mandela, chose to close down rather than to function under the racist regime. For decades after, South Africa offered a divided school system, with relatively well-functioning schools for the white minority, while the vast majority of black schools were characterised by poorly qualified and badly paid educators, dilapidated infrastructure, overcrowded classrooms, lack of textbooks and very basic learning facilities (Woolman & Fleisch, 2006). Government spending on black schools was one-tenth of the government spending on white schools (Byrnes, 1997; Buckland & Fielding, 1994; Fiske & Ladd, 2004; SAIRR, 2011). All of this had deep repercussions for many decades on the quality of education that black communities received.

The South African education system was also at the heart of the political struggle, and of the black student and teacher resistance movements against apartheid (Fiske & Ladd, 2004; Fleisch, 2002). The education sector was the scene of tragic historical events such as the Soweto uprising. In June 1976 more than 20,000 black students took the streets of Soweto to protest against the policy introduced by the Apartheid government to force Afrikaans as the medium of instruction in public schools. The clash between protesters and police lead to the tragic death of 200-700 young people (Gary, 2007; Harrison, 1987; Ndlovu, 2011).

With the ushering in of the democratic dispensation, the new government had to undo decades of inequality, restructure and create a new non-racial school system based on the values enshrined in the new constitution. Fourteen different education departments servicing the different black homelands (Bantustan States) as well as the white communities, where amalgamated into one national department of education with nine provincial counterparts (Bloch, 2009; Chisholm, 2004; Wolhuter, 2006).

The new Bill of Rights (Constitution of South Africa, 1996, Section 29) guaranteed every South African child, regardless of race and social-economic class, the right to basic education in any of the eleven official languages (Fiske & Ladd, 2004). Nine years of education was now compulsory for all children, regardless of their race and family income. In 1996 the South African School Act 84 (SASA) established the framework for a non-discriminatory inclusive education system, which also outlined new models of school governance and school financing.

In line with the spirit of the new democracy and the ANC's grassroots participation politics (Woolman & Fleisch, 2009), SASA encouraged the decentralisation of school governance. All schools were now led by autonomous School Governing Bodies (SGB), composed of principals and elected representatives of parents (who were the majority of the SGB), educators, administrative staff and learners (in secondary schools) (Chisholm, 2004; Fiske & Ladd, 2004; Grant Lewis & Motala, 2004). School Governing Bodies often include also other important community stakeholders such as local authorities, traditional leaders and local business and civil society. SASA was an important piece of legislation which provided parents a large degree of involvement and choice over the education for their children (Kanjee, 2012). The SGBs were granted significant powers such as developing a school constitution and mission statement, determining admission policy (as long as it was not discriminatory), the setting of the language of instruction and the religious observances, the administration of school property, the recommendation, promotion and appointment of staff, and the raising of additional resources through school fees (Fiske & Ladd, 2004; Woolman & Fleisch, 2006). Schools which would meet specific requirements of Section 21 of the SASA (1996) had further autonomy, were given the freedom to administer their own funds, recruit teachers, buy their own supplies and raise supplementary funds from external sources (families, businesses, communities, donors).

The SASA was accompanied in the same year by the National Education Policy Act 27 (NEPA), which set out further the responsibilities of the various spheres of education management in South Africa (Woolman & Flesich, 2009). In the new semi-federal government system, the 9 Provincial Departments of Education were responsible of implementation and delivery, while the National Government was responsible of setting education policy and providing funding (Motala et al., 2014). In 2009, the National Department of Education was later split into Department of Basic Education (DBE) and Department of Higher Education and Training (DHET) (Motala et al., 2014).

The NEPA further elaborated on the language policy in public schools, norms for school resourcing, and rules on the exceptions of school fees. NEPA also outlined the primary admission policy of schools to be set mainly on 'feeder zones' (geographic catchment areas), the size of which would be determined by Provincial Departments of Education (Woolman & Fleisch, 2006). Under the new policy, parents of children living or working within the area the School was located had first priority of admission (including the children of domestic workers) (Woolman & Fleisch, 2006). This facilitated some degree of redress of former inequalities, however as the school demographics reflected the population dynamics of the geographic locations (suburbs, townships, rural villages, etc.) in which they were situated, schools in South Africa ended up still to a large extent following the same social-economic lines of the previous Apartheid system (Motala et al., 2014).

Another key piece of legislation affecting the schooling system was the Employment of Educators Act 76 (EEA) of 1998, which governs the salaries, condition of service, recruitment, appointment, promotion, transfer, dismissal, retirement, rights and responsibilities of teachers in South Africa (Motala et al., 2014; Woolman & Fleisch, 2006). Roughly 95% of teachers in South Africa are employed by the provincial departments of education (OECD, 2008b). They are supported by the South African Council of Educators, the Educational Labour Relations Council (ELRC) and a number of trade unions, such as SADTU, NAPTOSA, SATU and PEU. Teachers are the single largest professional group in South Africa (DoE, 2008).

Many education policies have been implemented since 1994. Sayed et al. (2013) have accounted for over 160 policy texts; however, they have also argued that the South African bureaucracy and educator corps was not ready to implement some of these progressive policies in the appropriate manner. Some of the consequences of these various pieces of legislation and policy frameworks will be explored in subsequent sections.

3.3 Infrastructure, equity and access

South Africa has come a very long way in access to education since the 1970s, when the government was less concerned if black children were attending school or not. In the new democratic dispensation, access to schooling is mandatory for all children in the country, regardless of race, income or social class (Fiske & Ladd, 2004). Attendance to school is expected for all children aged between 7 and 15 (grade 1 to 9) and the government must ensure that no child is denied this important constitutional right (Motala et al., 2012). In recent years, the Government has also made efforts to include grade R (pre-primary) as part of the formal education system (see more in section 3.7.2).

As discussed in more detail in later chapters (chapter 4 and 5), to improve both access and quality of education services, both public and private actors have invested in many education infrastructure projects such as building, upgrading, expanding and refurbishing schools, classrooms, toilets, science labs, sporting facilities and libraries. While much efforts have been made to address equity and re-dress of previously disadvantaged communities, many schools, especially in the rural and poor parts of the country still lack adequate infrastructure, facilities, and resources to facilitate appropriate teaching and learning (DBSA, 2008). A DBSA (2008) report has highlighted that in South Africa 17% of schools still lack electricity, 61% lack appropriate water and sanitation, and 60 to 80 percent of schools lack other important learning facilities such as libraries, science and computer labs. Even when sophisticated facilities and upgrades to school infrastructure are provided by generous donors, these become difficult to maintain and sustain by the school and district authorities in the long-run.

The National School Nutrition Programme (NSNP) introduced since 1994 as part of the Government's Reconstruction and Development Programme (RDP), has also played an important role to promote not only increased access but also the alleviation of hunger and malnutrition in the poor schools and in rural areas (DBE, 2009). School feeding schemes are generally associated with the encouragement of enrolment of learners, however some international studies (Petrosino et al., 2012; White et al., 2012) have also found that correct and adequate nutrition does impact also on the cognitive abilities of children, and thus their ability to perform better also academically. Most of the poorest schools in South Africa are beneficiaries of the the state's school feeding programme (Taylor et al., 2013), and every day 9 million children receive at least one free meal at school paid by the government (Spaull, 2015). Attempts have been made in the past to conduct impact evaluations on the NSDP (Coetzee & Van der Berg, 2013), however this exercise has proved to be too challenging due to data limitations and the country-wide roll-out of the programme, which hinders the establishment of a proper counterfactual (see more in chapter 2).

In line with the commitments made at the Dakar World Education Forum for 'Education for All' (UNESCO, 2000), the post-1994 government has been committed and to a large extent successful in reaching MDG 2 of providing universal access to education in South Africa, with fairly good enrolment rates. Access to basic education in South Africa is near universal (98.6%) (Grant & Behrman, 2010). Enrolment to further education and training (Grade 10-12), which is not obligatory, is also fairly high 81% on national average (DBE, school statistics, 2013. Ninety-nine percent of learners in South Africa between 7 and 15 years old are registered in school (Taylor et al., 2013). Gender equity in the system is also well balanced with small differences existing in favour of girls. Children with disabilities are integrated in the school system with 90% participation rates (Croft, 2013).

Today the basic education system of South Africa covers 25,720 schools (including public, independent and special schools) that host 12,489,646 learners who are serviced by 425,023 educators

(DBE Education Statistics, 2014). Ninety-six percent of SA learners are the 'clients' in public schools (Bloch, 2009).

3.4 Curriculum frameworks

In South Africa's legislative transformations, curriculum was somewhat secondary to policies around finance, governance and organisation of schools (Christie, 1999). Nonetheless, the new government was faced with the challenge of radically transforming the previous racialist curriculum into a new balanced, unified and inclusive education programme which would provide the necessary training of young people who would serve the social, political and economic aspirations of the country (Harley & Wedekind, 2004; Fiske & Ladd, 2004). In 1995 the South African Qualification Authority (SAQA) Act provided the scaffolding of the national certification system and the academic and vocational training was merged together in the National Qualification Framework (NQF) established in 1996 (Chisholm, 2004; Fiske & Ladd, 2004).

Parallel to the democratic developments, South Africa saw a gradual introduction since the early 1990's of the philosophy of outcomes-based education (OBE) (Motala et al., 2014). OBE is a constructivist learner-centred open approach to education that sets out the general knowledge, skills and values learners need to gain, and allows educators the freedom and flexibility to design their own syllabus, teaching strategies, assessments and lesson plans in order to reach the required learning outcomes (Christie, 1999; Fiske & Ladd, 2004; Spady, 1994). OBE was very popular in Australia and New Zealand, and was introduced to South Africa in the early 1990s through the many visits of American consultant William Spady, who was a strong proponent of the approach (Jansen, 1999). In South Africa however it gathered its own 'local flavour' due to the empowerment and democratic values that were promoted during the period of political transition. Chisholm (2003, p.3) referred to OBE as the "the pedagogical route out of apartheid education".

In 1997 the Council of Provincial Education Ministers embraced OBE as the guiding principle of South Africa's basic education (Asmal, 2002), and the new Curriculum 2005 (C2005) was promulgated in South Africa's schooling system. Curriculum 2005 outlined 8 learning areas (Communication, Literacy & Language; Human and Social Sciences; Numeracy and Mathematical Sciences; Natural & Physical Sciences; Economic and Management Sciences; Technology; Culture, Arts and Artistic Craft; Life Orientation) for all 9 grades, totalling 66 outcomes for the whole General Education and Training (GET) phase of schooling (Chisholm, 2004).

Over the years OBE was a subject of great controversy and polarised opinions that sparked much academic and public debate around the appropriate approach for South Africa's education system (Bloch,

2009; Muller, 2001; Schollar, 2007; Spady, 2008). One of the earliest opponents was Jansen (1997, 1999) who criticized OBE to be overly complex, obscure, ridden with jargon and inaccessible to most South African teachers. Fiske & Ladd (2004) elaborated on the practical challenges of implementing the OBE curriculum within the South African schooling reality. The new curriculum was introduced too hastily in a top-down manner without having educators properly trained and prepared in the new pedagogical approach (Christie, 1999; de Clercq, 2008). OBE required a significant change of teaching culture (Fiske & Ladd, 2004), and a certain level of preparation of both teachers and students, which could be found only in historically white schools (Jansen, 1997). In the majority of the poorly resourced black schools with insufficiently trained and poorly supported teachers, the OBE experiment was a disaster (Fiske & Ladd, 2004)

The national curriculum was thus revised and simplified into the new National Curriculum Statement (NCS) in 2002 and later again in 2008 into the Revised National Curriculum Statement (RNCS) to align it to the National Senior Certificate (NSC) exam. In most recent years South African curriculum planners took a new approach which would focused more on teaching processes and inputs (see more in LTSM section) which was more digestible and more easily applicable to the South African schooling reality (Fleisch & Schöer, 2014; Schollar, 2015; Spaull, 2015). In 2011, the Department of Basic Education, approved a new Curriculum Assessment Policy Statement (CAPS), thus putting to a final end the era of OBE (Motala et al., 2014).

3.5 National learner assessments

Intrinsically connected to the development of any curriculum is the need of a system of assessments to monitor the delivery and implementation of the educational programme. UNESCO (2000, p. 14) defines learner assessments as "regular and systematic measurement exercises designed to determine what students have learned as a result of their educational experience". Assessment and tests have a function at learner level to assess competency, provide grading, certification, and degree of fit to labour markets (Muller, 2004). Learner assessments can also be used to assess specific education policy, programmes and the overall quality of the system of training and education (Kellaghan & Greaney, 2001). The National School Effectiveness Survey (Taylor et al., 2013) has also shown that monitoring and assessments are a critical factor contributing to learning outcomes.

In South Africa, periodic assessments have been important instruments for regular diagnostic of the education system and allow for the constant improvement of learning and teaching processes (Buhlungu, 2007). They assist in providing information and feedback to various school stakeholders (managers, teachers, parents, government) thus increasing dialogue, mutual accountability, identifying

challenges, and adjusting education policy, instruction methods and school systems, to achieve better learning outcomes (DBE, 2011; World Bank, 2011). In the education literature this is often referred to as internal accountability (Elmore, 2008) or school-based accountability (Taylor, 2009).

In the previous chapter 2, we have discussed how the terms 'formative' and 'summative' evaluation (Scriven, 1991), have originated in the education literature, to indicate assessment done internally with the purpose of learning and improvement in the case of formative, while summative referring to assessments done at the end of the learning process by external independent bodies for the purpose of judging fitness or achievement of certain yardsticks. Kanjee (2007) has distinguished several types of assessments done in education: school-based, provincial/district, national, international, small assessments for the the purpose of evaluating programmes, and public examinations.

Post-1994, South Africa has had many efforts – school-based, provincial, national and international (discussed more in next section) – to monitor the quality of the education system (Spaull, 2013). As a subsection to the NEPA (1996), the Assessment Policy in General Education and Training: Grade R to 9 and ABET was formulated by the Department of Education in 1998. Assessment guidelines for inclusion in Curriculum 2005 (DoE, 2002) and an interim policy framework for the assessment and promotion of learners in Grade 9 (DoE, 2003) were later expanded. A national protocol for assessment for schools in the GET and FET bands (Grade R to 12) was developed to standardise reporting within the framework of the National Curriculum Statements (DoE, 2005). The latest assessment policies (DoE, 2007) placed greater emphasis on classroom tests by outlining the range of assessment methods and the types information required by the teachers and other stakeholders, providing structure and templates for reporting (Kanjee in Buhlungu et al., 2007).

The primary public examination that occurs in South Africa at the end of the formal schooling process (Grade 12) to signal the individual's readiness for tertiary education is the Senior Certificate (SC), commonly referred to as 'matric' exam (Kanjee, 2007). Since 1958, South Africa has had a history of high school exit examination (matriculation), which has undergone several permutations (Joint Matriculation Board, 1918-1992, SAFCERT, 1992-2001), until it reached the current format of the National Senior Cetificate (NSC). In 2001 the General and Further Education and Training Quality Assurance Act, established Umalusi as an independent statutory body, which would manage the national learning assessments and provide certification for learners emerging out of the general and further education and training band (grade R-12), based on their results on the NSC exams. The NSC is a standard examination administered across the country and marked and moderated centrally by Umalusi for all learners completing grade 12. It is aligned to the National Curriculum Statements and provides an indication on every learner's capacity to enter and pursue higher education and training. It has been modified a few times, but in the current format learners need to achieve 30% or above in the language of

learning for a Certificate pass, 40% and above in 4 additional approved subjects for a Diploma pass, and 50% and above in 4 additional approved subjects for a Bachelors pass, which would indicate a good degree of learner preparedness for undergraduate University studies.

The NSC has also been an important instrument to signal the quality of South Africa's education system (Reddy, 2006), and used by the DoE to keep schools accountable for their performance (Kanjee, 2007). In South Africa, the NSC pass rates and marks have also been used extensively in education research and evaluation (see more in chapter 5 and 6), as they provide a good indicator of learner achievement. Matric results are available on an annual basis for all schools and all learners in the country, and data is carefully reviewed for quality assurance by Umalusi. Aside from the initial years (1999-2003) where the marking process was questionable and a number of administrative breaches occurred (Taylor, 2009; Umalusi, 2004), the NSC data has been fairly consistent since 2004 (Muller, 2004).

In 2008, the National Curriculum changes did affect also the NSC examination, which saw some changes in the subjects and grading system. Mathematics became obligatory but was split in pure mathematics (previously Higher Grade (HG) and mathematics literacy (previous Standard Grade). Since the introduction of mathematics literacy more learners opted to take this subject as it was easier to pass (Spaull, 2013). According to Simkins (2010) the 2008 revisions in the NSC saw a grade inflation and made it easier for learners to pass the subjects of mathematics and sciences. Hunt, Schoer, Ntuli, Rankin, and Sebastiao (2011) have also shown that with the introduction of the new system, there has been a general down-grading of NSC marks.

Several South African education experts (Fleisch, 2006; Jansen, 2004; Kanjee, 2007; Spaull, 2013) have argued that NSC results cannot be taken at face-value as many learners drop out of school in earlier grades and teachers and managers deliberately hold back unprepared learners in grade 11 to reduce the failure rates in grade 12, thus providing an inaccurate picture on learner achievement in South Africa's education system. Another critique of the NSC examinations is that it puts a lot of emphasis in one particular summative moment at the end of the schooling process, thus encouraging more interventions to occur at later educational phases to the detriment of important programmes for earlier grades (Motala et al., 2014), where international studies (Black, & Lipsey, 2008; Bloom et al, 2008) have proven that most learning gains occur. Efforts have thus subsequently been made to expand national learner assessments also to earlier grades of South Africa's public system.

Among the early assessment policies of the Department of Education (DoE, 2001), was the introduction of the framework for 'Systemic Evaluations' which tested a representative sample of approximately 54,000 learners of grade 3, 6 and 9 in more than 2000 primary schools of the country (DoE, 2008; Spaull, 2013). The systemic evaluations consisted of literacy, numeracy and life skills tests, accompanied by questionnaires to gather more contextual information on the learners, teachers, principal

and parents, which might affect performance (DoE, 2001). In later years, a more systematic approach was taken through the National School Effectiveness Study (NSES), which used longitudinal panel data to test the same pupils in a sample of 266 school in grade 3 (2007), grade 4 (2008) and grade 5 (2009) (Taylor et al., 2011). At the provincial level standardised learner assessments have been conducted in the Western Cape Learner Assessment Study, which was conducted in grade 6 of every primary school of the province (WCED, 2005).

Until recently the only nationwide standardised yearly examination was the NSC, which occurred at the end of the education system (grade 12). All other national and provincial assessments discussed above, as well as the international tests implemented in SA (discussed more in the next sections), were normally done with only a representative sample of schools, classes and learners. Consequently, there was no standardised assessment done in primary schools across the country, which would allow for comparison of schools and provinces, and allow for accountability for underperformance (Spaull, 2013). In recent years, the importance of regular school assessments at earlier grades has been highlighted by the Department of Performance, Monitoring and Evaluation (DPME, 2010) as well as by the Department of Basic Education in its 2011-2014 Action Plan (DBE, 2011). The Delivery Agreement between the President and the Minister of Education (2010) states that Annual National Assessments (ANA) will become the 'cornerstone of the quality improvements in South African schools'.

ANA consists of nationally-implemented standardised literacy and numeracy tests for all learner in grade 1 to 9. DBE provides question papers and marking memoranda (exemplars) and schools manage the conduct of the test as well as the marking and internal moderation (DBE, 2015). ANA serves mainly as a classroom improvement tool for teachers to assess learner knowledge and report back to principals and parents through learner report cards. The first round of Annual National Assessments (ANA) was implemented in 2011. In that year, HSRC performed also a 'verification ANA', where a sample of 50 test scripts from grade 3 and grade 6 from 1800 schools was remarked and checked (DBE, 2011). The ANA verification was also accompanied by supplementary questionnaires for learners, educators and school principals, to better understand the various in and out of school factors contributing to the learning results. In subsequent years, verifications were done directly by the DBE, by remarking a sample of scripts from grade 3, 6 and 9 (DBE, 2012).

ANA promises to be a very useful tool for systems improvement, internal accountability and to assist with research and policy in South Africa's education sector. Nic Spaull (2013, p. 14) has described ANA as a hallmark achievement of the DBE and the "largest data-gathering exercise in the country apart of the two censuses". The assessment instrument however comes also with a lot of limitations and criticism. Taylor (2011) has warned that ANA does not provide a reliable and consistent standard across schools as it is administered, invigilated, scored and moderated differently by the school teachers,

principals and district officials, who will be affected (positively or negatively) by the results of the assessment, and thus strongly prone to manipulation and gaming (Fleisch, 2006). Van der Berg and Spaull (M&G, 2012) have elaborated on how ANA cannot be appropriately used for inter-school comparison, and not even for time-series comparison, as the difficulty levels of the ANA tests have changed significantly over the years.

Aside from the above learner assessments administered by the National and Provincial Departments of Education, a number of specialised education think-tanks, non-governmental organisations (Muller, 2004) and private consultancy firms, such as Eric Schollar and Associates (ESA), Joint Education Trust (JET) and the Human Sciences Research Council (HSRC), have been designing and administering their own learner assessments, which are often done to evaluate specific programmes on behalf of private funders or governmental donors. The use of these evaluation-related tests versus state-run standardised assessments will be discussed further in chapter 5 meta-analysis. In conclusion, Graeme Bloch (2009) has reminded us that all these learner assessments focus only on academic abilities and do not say much about values of citizenship and social responsibility, which also characterise South Africa's schooling curriculum. What is often reported about these test are averages of schools and provinces, which still hides much of the inequality and extreme differences in performance still present within the South African education system (Bloch, 2009).

3.6 The 'crisis' of quality

As discussed earlier the focus of the Millennium Development Goals and the UNESCO 'Education for All' (2000) campaign has been on extending universal access, however quality of education systems has a more critical role in ensuring growth in developing countries (Hanushek & Woessmann, 2012). "Once children are through the school gates, it should be expected that their education will provide them with at least a minimal set of skills, knowledge and values that will allow them to function in the economy, in social life and in the democratic process" (Motalaet al., 2012, p.2). Evidence over the past decade has shown that efforts to expand enrolment must be accompanied by attempts to enhance educational quality if children are to be attracted to school, stay there and achieve meaningful learning outcomes (UNESCO, Bangkok, 2010).

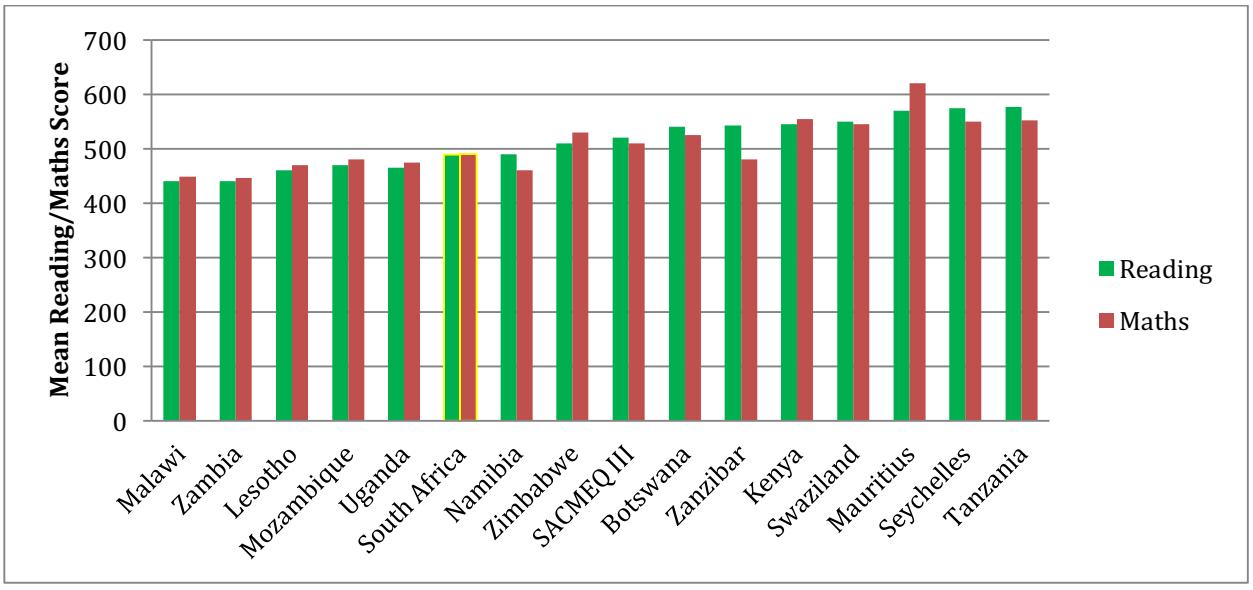
In the post-1994 dispensation, many efforts have been implemented to address inequality in the access to education, but inequality in the quality of education in South Africa still persists, and this has been an ongoing concern of the ANC government (Sayed et al., 2013). Education scholars (Bloch, 2009; Motala et al., 2012; Motala et al., 2014; Taylor et al., 2013) have acknowledged that South Africa has effectively a two-tier public school system: a small group of well-functioning schools catering for the

well-off elites, and a large majority of dysfunctional schools servicing the black, rural and economically disadvantaged communities. South Africa has the highest levels of between-school performance inequality in both mathematics and reading in comparison to other neighbouring countries like Botswana, Lesotho and Swaziland (Motala, 2006; Van der Berg, 2011). Only twenty percent of South African schools (including private schools, former model C, and few well-performing black schools) are producing the university graduates of the country (Bloch, 2009).

A range of commentators (Bloch, 2009; Fleisch, 2008; Spaull, 2013) have discussed the 'crisis' engulfing South Africa's education system. A recent OECD report has ranked South Africa number 75 out of 76 countries with regard to the quality of education (OECD, 2015). These signals are systematically confirmed by a number of international education surveys, where South Africa consistently scores lower than other middle-income countries and even below some of its poorer African neighbours (Strauss & Burger, 2000; Spaull, 2013). One of the oldest of these international assessments, the Monitoring Learning Achievement (MLA) run by UNESCO/UNICEF in 1999, had ranked South Africa, with and average national score of 30% (Bloch, 2009), in the bottom 12 African countries in numeracy. In the 2003 Trends in International Maths & Science Study (TIMSS), as well as the 2006 Progress in International Reading Literacy Study (PIRLS), South Africa's average scores were below countries with lower income such as Botswana, Morocco, Chile and Malaysia (Spaull, 2013).

A more thorough assessment of both numeracy and literacy at grade 6 was done through the various surveys of the Southern African Consortium for Monitoring Educational Quality (SACMEQ). In SACMEQ assessments, South Africa performs below the African average and below countries such as Swaziland, Tanzania, Botswana and Kenya (Spaull, 2011). SAQMEQ data shows that in grade 6 tests, 52% of South African mathematics learners and 31% of language learners had competencies comparable to grade 3 learners or below (Figure 16). The 2010 SACMEQ study show that South Africa has not made much progress from the last assessment (2005) and that very large differences still remain in learning results between different schools in the country. The SACMEQ survey (2010) has also highlighted that high repetition rates have been a major factor in decreasing the quality of education in South Africa.

Figure 16. Mean reading and maths scores (SACMEQ III).



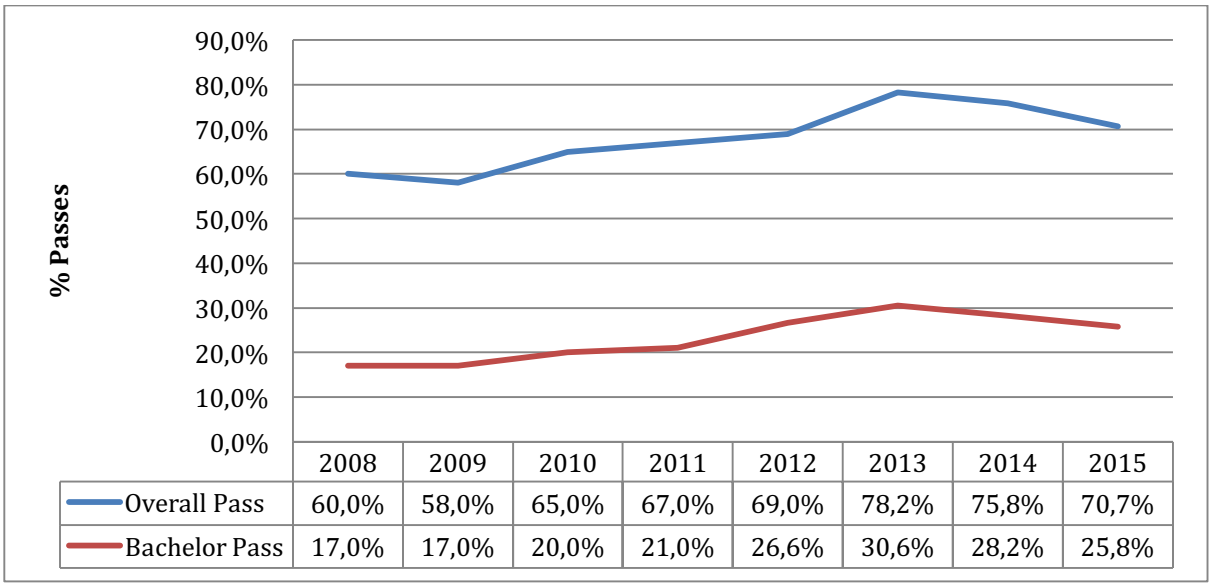
Source: N Spaull, 2011. A Preliminary Analysis of SACMEQ III South Africa

The Global Competitiveness Report ranks South Africa 127th among 142 countries with regard to quality of primary education (World Economic Forum, 2011). The Education Roadmap Diagnostic (2008) explains that only 10% of SA learners compare to their counterparts in developed countries with regard to literacy and only 6% in the areas of mathematics and sciences.

Numerous other studies done nationally (Howie et al., 2007; Pereira & du Toit, 2012; Reddy et al., 2006) point to the same alarming signs on South African learner achievement. The early Systematic Evaluations found that only 35% of children passed the language tests, 27% passed mathematics, and 41% passed natural sciences tests (Bloch, 2009; DBE, 2005). The HSRC-administered Annual National Assessments (ANA) in 2011 showed that in grade 3 the average literacy and numeracy scores were 35% and 28%, respectively, while in grade 6 they were 28% and 30%, respectively (DBE, 2011). Spaull (2015) has emphasised that half of grade 4 children in South Africa cannot understand what they read and 76% of grade 9 learners are 3-4 years behind in terms of their mathematics knowledge.

The trends on the NSC examinations show that over the years pass rates have gradually improved. In 2015, of all learners who registered for the NSC, 16.4% passed with higher certificate, 28.5% with diploma level and 25.8% with bachelor pass (see Figure 17).

Figure 17. Grade 12 pass rates and Bachelor passes.

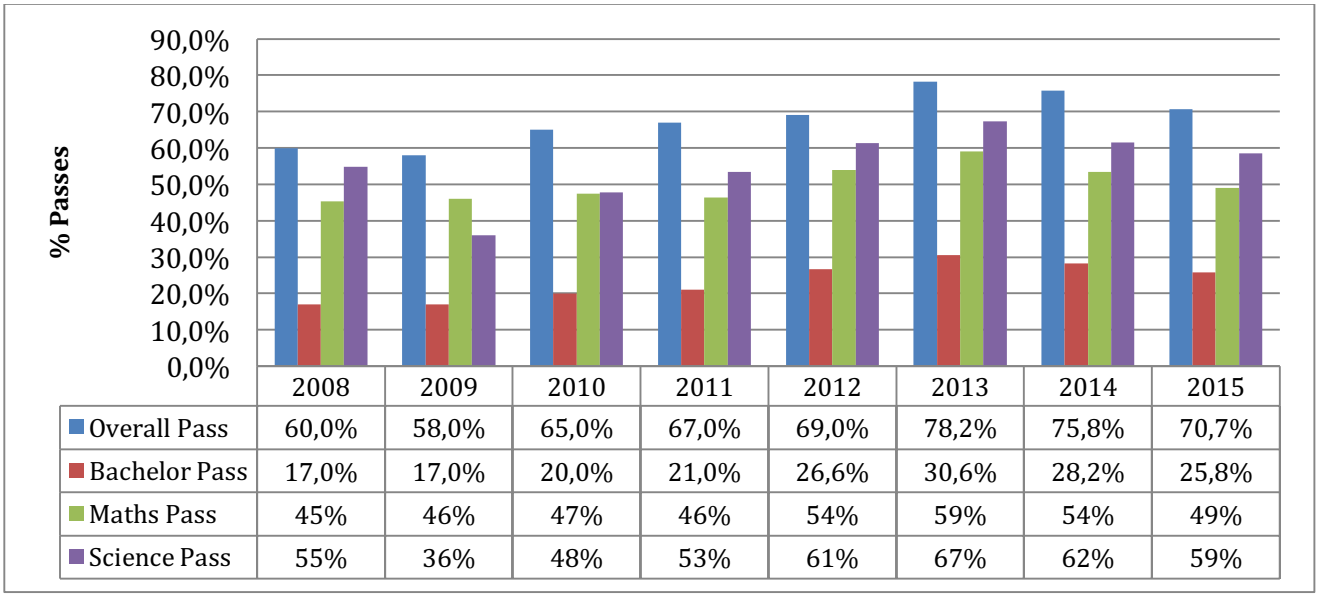


	2008	2009	2010	2011	2012	2013	2014	2015
Overall Pass	60,0%	58,0%	65,0%	67,0%	69,0%	78,2%	75,8%	70,7%
Bachelor Pass	17,0%	17,0%	20,0%	21,0%	26,6%	30,6%	28,2%	25,8%

Source: DBE 2011-2015, NSC Examination Report

The results however are masked by the fact that FET schools (formerly known as TVET colleges) have dramatic drop out rates and not many students reach the final year, where they register for the NSC examination. Generally speaking, schools across the country contain half the number of learners in Grade 12 than they have in Grade 10 (DBE, 2012, see Figure 18). Half of those who exit the schooling system do so post grade 11, either by not enrolling in grade 12 or failing the NSC (NPC, 2011). On a number of occasions, Nic Spaull (2013, p.10) has vividly illustrated that "of 100 pupils that start school, only 50 will make it to Grade 12, 40 will pass, and only 12 will qualify for university" and only 5% will get a degree (Spaull, 2015).

Figure 18. NSC results for maths and science subjects.

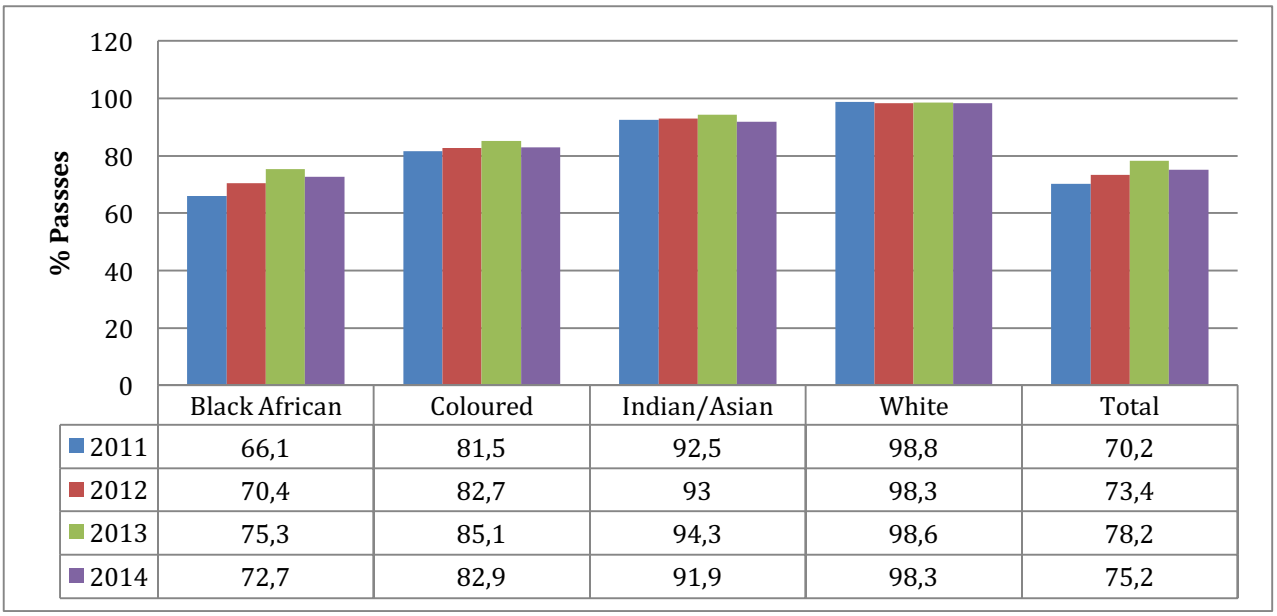


	2008	2009	2010	2011	2012	2013	2014	2015
Overall Pass	60,0%	58,0%	65,0%	67,0%	69,0%	78,2%	75,8%	70,7%
Bachelor Pass	17,0%	17,0%	20,0%	21,0%	26,6%	30,6%	28,2%	25,8%
Maths Pass	45%	46%	47%	46%	54%	59%	54%	49%
Science Pass	55%	36%	48%	53%	61%	67%	62%	59%

Source: DBE 2011 and 2015, NSC Examination Report

The NSC also confirms South Africa's poor performance in mathematics and science subjects. In his work, Schollar (2008) estimated that of the total number of learners enrolling in grade 1 in the post-1994 schooling dispensation only 1,5% achieved a Mathematics HG pass during their 2006 grade 12 Senior Certificate (SC) examinations. To take a more practical case, in 2003 two townships, Tsakane and KwaThema, situated in Gauteng, the heart of the nation's economy, had together 1600 matric passes but only 12 of these were in HG mathematics. In 2015 of all learners who took mathematics and sciences only 54% passed above the 30% mark, and 31.9% passed above the 40% mark (DBE, 2016). These poor trends in mathematics and sciences are accentuated even more in the case of black students (NPC, 2011). This hinders previously disadvantaged populations from accessing much of the higher education opportunities, and in turn very few black learners become engineers, scientists, financial analysts, and other technical professions critical for the South African economy. The Centre for Development Enterprise (CDE, 2010a) has identified the shortage of black graduates in mathematics and sciences as the biggest obstacle to economic growth, equity and black economic empowerment in South Africa.

Figure 19. NSC pass rates by population group.



	Black African	Coloured	Indian/Asian	White	Total
2011	66,1	81,5	92,5	98,8	70,2
2012	70,4	82,7	93	98,3	73,4
2013	75,3	85,1	94,3	98,6	78,2
2014	72,7	82,9	91,9	98,3	75,2

Source: Vilakazi/DBE, 2016- Presentation on NSC achievement by racial groups

As discussed throughout this section, wide disparities still exist within South Africa's education system, with the quality of pass rates still varying greatly depending on race and social-economic status (Vilakazi, 2016, see Figure 19). Spaull (2015) has argued that even 20 years into democracy South African previously disadvantaged learners are still caught in the same cycle of poverty because of the low quality of the education system:

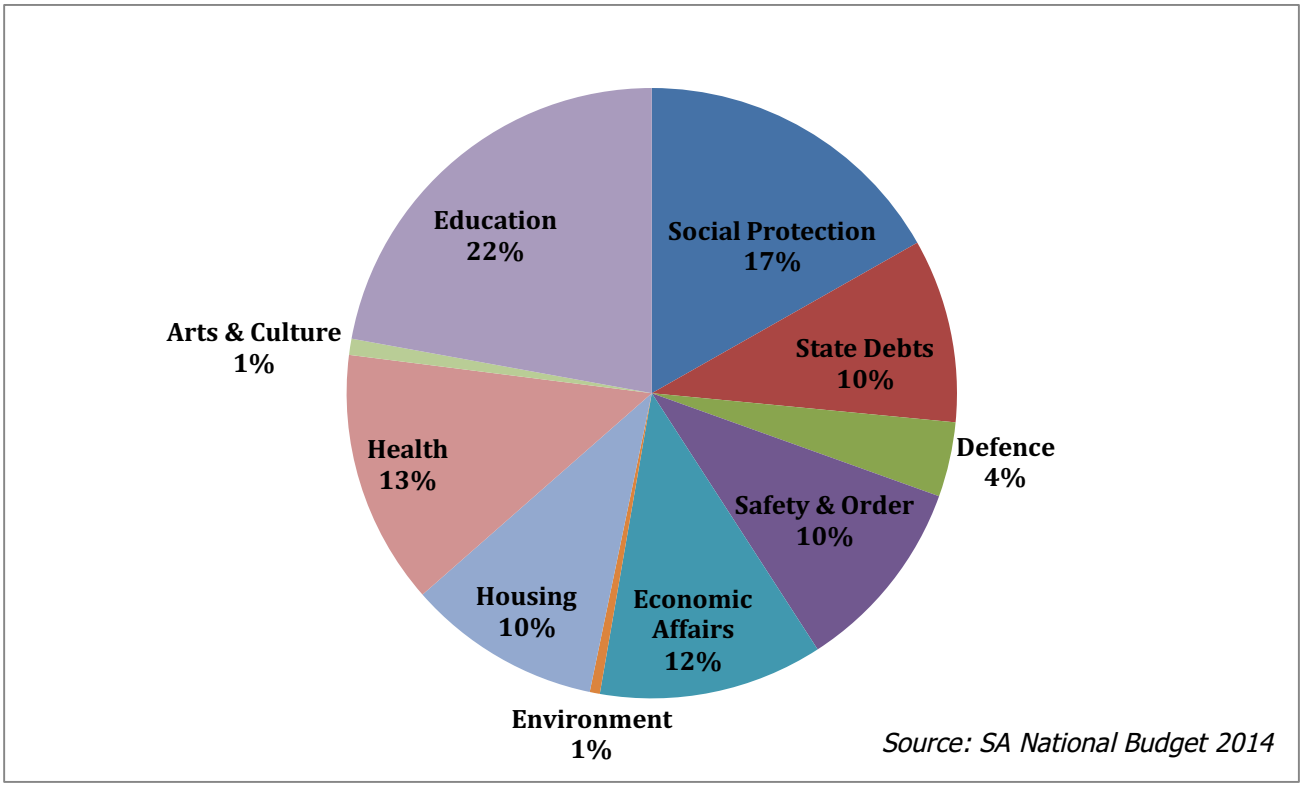
"I maintain that we have an ongoing crisis in education and that poor children continue to be condemned to hereditary poverty as a direct result of the low quality education they receive at school. Poor quality education was and is a poverty trap. This should be our biggest source of national shame" (Spaull, 2015, education blog).

3.7 Funding education development in South Africa

3.7.1 Government Spending

In the post-apartheid dispensation, education has occupied a high profile in government policy and campaigns. In order to promote sustainable economic growth, redress past inequalities and promote African empowerment, education has been seen as a vital tool to achieve the transformational agenda of the ANC government. Of the 12 National Outcomes of President Jacob Zuma, 'Quality Education for All' features as the first one. In its Diagnostic, the National Planning Commission (NPC, 2011) has identified high unemployment and poor educational outcomes as the two most pressing issues facing the country. Education, innovation, and training are also a major component of the NPC's Vision 2030 National Development Plan (NPC, 2012). As a result, education receives the biggest share of public spending in South Africa (National Treasury, National Budget Statements, 2010-2015, see Figure 20).

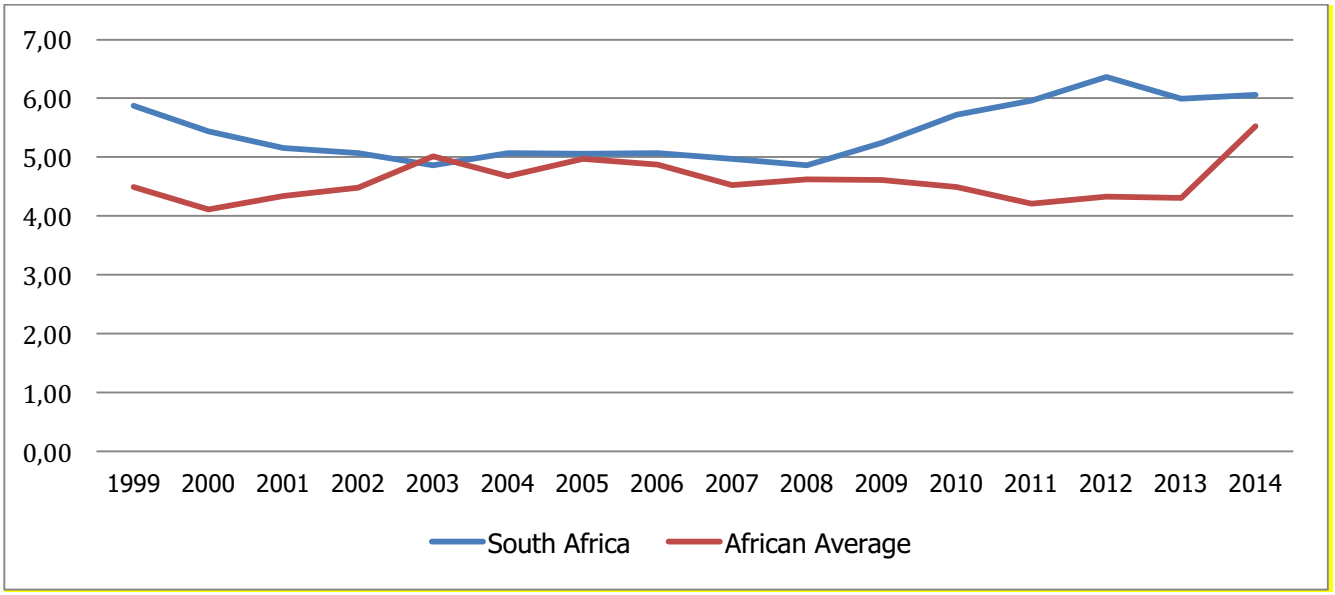
Figure 20. Percentage of SA national budget allocations by sector.



In its 2015/2016 budget the government allocated R276.7 billion to education occupying 22% of government spending. While all provinces receive a specified budgetary allocation based on a predetermined demographic formula (PFMA 1999) (National Treasury, 2016), provincial legislators can allocate different amounts of the provincial government budget to education (Fiske & Ladd, 2004). As a result, each of the nine Provinces of South Africa allocates between 30 and 50 percent of their annual budget to educational services (National Treasury, ENE, Provincial Budgets, 2015).

National and provincial departments recurring budget lines include the following: administration; curriculum policy, support and monitoring; teachers, human resources and institutional development; planning, information and assessments; and educational enrichment services (National Treasury, ENE, 2015). Personnel costs occupy 80-95% of the educational expenses according to different estimates (OECD, 2008b, World Bank, 2011); however, significant investments are also made on facilities, school meals, textbooks, and other learning materials. To address the major infrastructure deficit in poor schools the government has also instituted a number of conditional grants such as the Accelerated Schools Infrastructure Delivery Initiative (ASIDI).

Figure 21. Percentage of GDP spent on education



Source: UIS Statistics 2015

For the past decade, public spending on education has represented 5-6% of the country's GDP (Figure 21). By global standards this is fairly high and comparable to other countries renowned for their high education spending such as Brazil and Australia. South Africa also meets the UNESCO global benchmark of 6% of nation's GDP in education spending. In other developing countries and in sub-Saharan Africa the average is 3% (AAI, 2015). Teachers in South Africa are among the highest paid in the world in purchasing power terms (NPC, 2011), and every learner in South Africa receives on average $1383 compared to the average education spending in the rest of sub-Saharan Africa with $167 per learner (DBE, 2012; Taylor, van der Berg & Mabogoane, 2013).

Since the South African Schools Act (1996), pro-poor funding policies were introduced to address historical inequalities between white and black schools and promote redress (Woolman & Fleisch, 2006). The National Norms and Standards for School Funding (DBE, 2014) favour the rural and economically disadvantaged schools. In the new dispensation schools were classified into five quintiles based on a poverty index, that took into account average household income, unemployment rates, and levels of education of the community around the school (Sayed & Motala, 2012). Schools in different quintiles would receive different amounts of public funding, with the lowest quintile schools receiving the greatest amount of non-personnel resources per capita for each learner (Woolman & Fleisch, 2006). As a result, provinces with large number of quintile 1 and 2 schools (i.e., Limpopo and the Eastern Cape) receive more public funding than provinces where the majority of the schools are in the higher quintiles (i.e., Western Cape) (Sayed & Motala, 2012). Between 1991 and 2005 public spending on African learners increased by 75% (Gustafsson, 2007) and the gap between what learners receive overall in the richest quintile schools and the poorest quintile schools (R5,284 and R4,757, respectively) has been gradually decreasing (OECD, 2008b, see Table 2).

Table 2. *National Expenditure Per Quintile*

		2007		2008		2009	
Quintile	% Learners in the Quintile	Rand/Child		Rands/Child		Rands/Child	
NQ1	30%	R	738.00	R	775.00	R	807.00
NQ2	28%	R	677.00	R	711.00	R	740.00
NQ3	23%	R	554.00	R	581.00	R	605.00
NQ4	15%	R	369.00	R	388.00	R	404.00
NQ5	5%	R	123.00	R	129.00	R	134.00
Overall	100%	R	492.00	R	517.00	R	538.00

Source: DoE 2010

The National Norms and Standards for School Funding (NNSSF) were amended in 2006 so to allow quintile 1 and 2 to be fee exempt (World Bank, 2011). The no-fees policy was extended in later years to also include quintile 3 schools (DBE, 2011), allowing thus for 70% of learners in the country to receive effectively free education paid by the State (DBE, 2010, Sayed & Motala, 2012).

3.7.2 Parental Spending

The South African Schools Act (SASA) (1996) makes provision for independent (private) schools to operate in South Africa within the frameworks and policies of the Department of Education. Section 39 of SASA also encourages public schools to supplement their income by charging fees and seeking additional private funding (Motala et al., 2012). While quintile 1, 2 and 3 operate on a no-fees basis (see previous section), quintile 4 and 5 public schools as well as independent schools in South Africa charge user fees that allow schools to raise more money to hire more and better quality teachers. In some cases, such as the Ermelo former model C school, half of educators are state-funded and half are on contracts provided by the SGB (Fleisch, 2016).

Woolman and Fleisch (2006) have argued that the policy framework created by the SASA (1996), the NEPA (1996) and the EEA (1998) has developed the incentives for parents, principals, school governing bodies, to give rise to a 'quasi-market' approach to South Africa's schooling sector. Ahmed

and Sayed (2009) have also echoed that the open policy on school fees has created a semi-privatisation system of the education sector. The constitution of South Africa, along with the subsequent education legislation introduced in the post-1994 era, has created a combination of utilitarian and quasi-libertarian basic education system in South Africa (Woolman & Fleisch, 2009), allowing the state to address the needs of the poorest, while keeping the privileged in the system, by providing them opportunities to improve the quality of their children's education.

As it has internationally, in South Africa the policy of school fees - closely linked to issues of access, adequacy and equality (Roithmayr, 2002; Woolman & Fleisch, 2008) - has caused much debate and controversy among many education experts (Fiske & Ladd, 2004; Lewis, & Motala, 2003; Vierava & Wilson, 2005). Some argue that school fees promote the privatisation of the education system and continue to foster the Apartheid era racial and class inequalities (de Groof & Bray, 1996; Fiske & Ladd, 2004; Pampallis, 2003; Roithmayr, 2003). Others explain that school fees have led to more local control of resources, better efficiency and delivery, creating a healthy competition among schools to improve quantity and quality of state services, so to appeases the rich families and keeping them in the system, while freeing up resources for the weaker schools, and promoting cross-subsidisation of poor learners by rich learners (Sayed & Motala, 2012; Woolman & Fleisch, 2006)

Regardless of the political and academic views, the fact remains that school fees provide over 3,5 billion rand a year of income to schools that supplement the funding they receive from the government (Fleisch & Woolman, 2004). In discussions around private financing for education development in South Africa, parents' contribution is a bigger force than income from traditional donors and the business sector, which will be discussed more in the next section.

3.7.3 Private sector investments

The introductory chapter of this thesis has alluded to the role of emerging actors, particularly the private sector, as important partners in the achievement of the Sustainable Development Goals (SDGs) and in addressing continental and national development challenges (Besharati, 2013c). South Africa presents a good example where corporate donors have contributed substantive resources to support the public sector's efforts to improve the country's schooling outcomes. The following section on the private sector's investments in South Africa's education sector, is based on primary research conducted by the author in this field, findings of which have been published by the Joint Education Trust (Besharati, 2015) and the South African Institute of International Affairs (Besharati 2012, 2014).

South Africa's interesting history of private sector's engagement in education, has deep roots and is linked to the role education plays in South Africa's social and political transformation (Besharati,

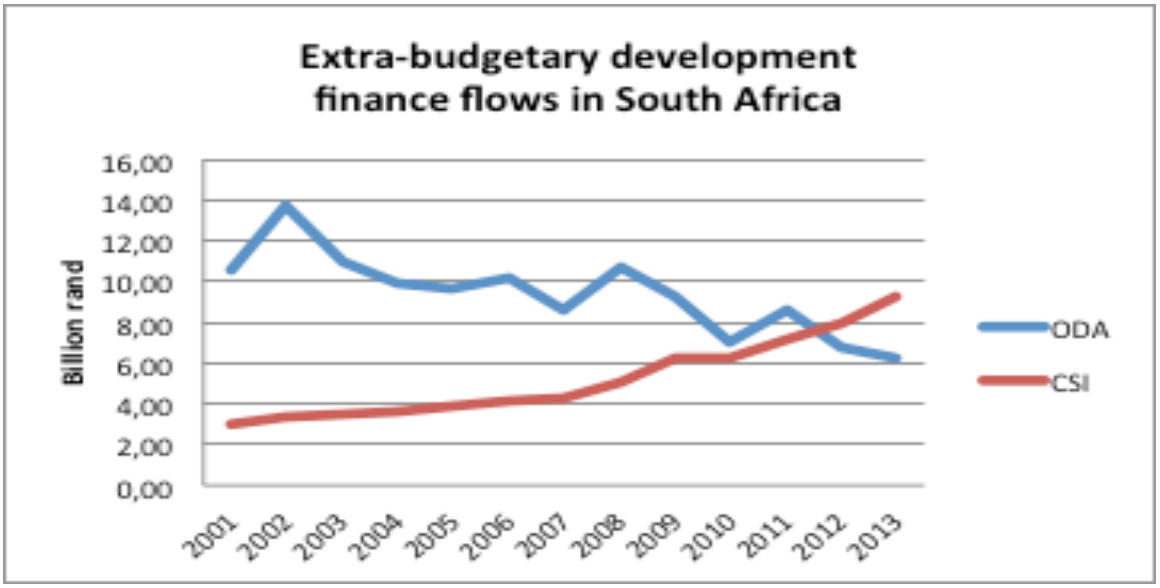
2015). The nation's economy is highly dependent on the availability of skilled labour, which is raised through a well-functioning school system. This provides a strong 'business case' for private sector to partner with government and provide all kinds of support and development programmes to improve the quality of education in the country (Besharati, 2014). Nonetheless, corporate motivations to engage in the education sector have been of mixed nature. Though there is some element of altruism and philanthropy in private giving, CSI are also public relations 'window-dressing' exercises, marketing efforts to promote a good responsible image of the corporation (Besharati, 2015). Increasingly, a new approach of 'enlightened self-interest' (Whittaker, 20 May 2013, personal interview) is emerging, where, companies invest in education, knowing that uplifting human capital in South Africa will also accrue long-term returns to their businesses. As will be illustrated in more detail in the Anglo American case study in chapter 4, corporations have been concerned with the development of a young, well-educated and capable African population, to improve workforce productivity, and increase the pool of quality service providers and clientele to engage with (Besharati, 2012, 2014). These are all critical for improving the country's investment environment and international competitiveness. As American politician and businessman Charles E. Wilson (1957) said 'what is good for America is good for General Motors and vice versa', South African businessman and social investor Brian Whittaker (Whittaker, 20 May 2013, personal interview) echoed 'what is good for South Africa is also good for Anglo American' (Besharati, 2015). Public and private interest converge and support one another when there is a stable nation, and a prosperous and growing economy, which in turn leads to human and social development.

In South Africa a series of important policy frameworks have provided incentives for corporations to invest more in public goods and in the social arenas (Besharati, 2015). These have included, tax deductibility on donations to PBOs (TLAA, 2000), the Broad-Based Black Economic Empowerment (BBBEE) Act (2003), which paved the way to BBBEE scorecards and industry charters, the Mervyn King Reports on Corporate Governance, the initiation of the Socially Responsible Investment (SRI) Index (2004) in the Johannesburg Stock Exchange (JSE), and the Mining and Petroleum Act (2002), which requires the extractive industries to develop with local stakeholders Social and Labour Plans (SLPs) (2002), that allow companies to receive licences and concessions based on their commitments to capacity-building and community development (see more in chapter 4).

As CSI in South Africa are scattered and fragmented across a range of different players, it is difficult to measure the full extent of social spending, nonetheless education ranks consistently at the top of the list of the sectors supported by companies (Perold & Associates, 2012; Trialogue, 2015). The Cape Town based consulting firm, Trialogue, conducts a yearly survey of a sample of the biggest South African and international companies and publishes such information in its yearly CSI Handbooks. In 2014/2015 financial report, CSI rose to over R8.1 billion (Trialogue, 2015). Between 2005 and 2013

education spending constituted between 35 to 45 percent of CSI in South Africa, against for instance funding to health or environment sectors, which have been below 20% (Trialogue, 2005-2013). In another survey commissioned by National Business Initiative (Perold & Associates, 2012), corporations were found to be contributing over R1.3 billion to different education programmes, from pre-primary to tertiary and adult education. These figures are probably under-estimated, as CSI budgets do not usually include company products, services and employee time, which normally is included in traditional donor's aid budgets (OECD, 2011). Both Trialogue and NBI survey mainly the large businesses and do not include the small and medium enterprises, which Perold & Associates (2012) estimate to contribute an additional R4 billion a year to education. All trends however indicate that the investments made by the corporate sector of South Africa's surpass by and large aid from traditional donors to South Africa's education sector, which in 2011 and 2012 was reported to be merely between R150 to R300 million (National Treasury, 2012). A company such as Anglo American has a yearly CSI budget which surpasses the ODA (Figure 22) budget of the likes of DFID, the World Bank, the United Nations, GIZ and France (Besharati, 2014).

Figure 22. Corporate social investments versus foreign aid to SA



Sources: Authors own compilation based on data from OECD, National Treasury & Trialogue

Corporate sector engagement in the education sector of South Africa has emerged out of a long history of public-private partnerships where business leaders would pool resources together to support development

objectives of the country (Besharati, 2015). In response to the tragic Soweto riots of 1976 (see earlier section), 180 business leaders gathered for a conference at the Carlton Hotel to address the plight of the disfranchised black population. This conference led to the establishment of the Urban Foundation (UF), that pooled resources from the 80 corporations in order to provide housing and schools for poor communities (Pillay, Tomlinson, & Du Toit, 2006). In the early 1990s, 14 businesses established a partnership with the trade unions (COSATU, NACTU, SADTU) and the black political parties (ANC, IFP, OSAPO, PAC) called the Joint Education Trust (JET) to address the challenge of restructuring the country's dire education sector in the post-apartheid era (JET, 2016). Since 1992 JET has been an important vehicle for domestic and international donors to channel over R1 billion to hundreds of non-profit service providers (Fiske & Ladd, 2004) to implement educational programmes, provide policy advice, and conduct education research to test and refine school development models.

One of the first education public-private partnerships of post-1994 South Africa was the flagship programmes of the National Business Initiative (NBI), the Education Quality Improvement Programme (EQUIP), which was implemented in 500 schools throughout the country (Sayed et al., 2013). The evaluation of the EQUIP will be further discussed in more detail in chapter 5. One of the biggest partnerships implemented in South Africa was the Business Trust, which saw on its executive board a large number of senior government officials and business leaders of the country (Business Trust, 2014). Among its manifold programmes, the Business Trust invested R400 million between 1999 and 2005 to improve hundreds of schools across South Africa, through its Quality of Learning Project and Learning for Living (see evaluations of both programmes in chapter 5). A more recent public-private partnership is the National Education Collaboration Trust (NECT), which is one of the most prominent tripartite accords between social partners[1] (MetCalfe, 16 February 2013, personal interview) for the implementation of South Africa's National Development Plan (2012).

In South Africa companies engage in education activities through some of the platforms illustrated above, but also independently through their own CSI divisions, their trusts and foundations, or through fund managing firms such as Tshikululu Social Investment (TSI). The majority (70%) of private sector funding for development is channelled through NGOs, NPOs, charities and affiliated schools (Besharati, 2015). In the case of the manufacturing sector and the extractive industries (Mining and Petroleum Act, 2002), companies are encouraged to service first the communities (hence also the schools) around their geographic area of operation and production (see more in chapter 4).

[1] Government, business, labor unions and communities

Table 3. *Largest Corporate Education Investors in SA*

Company	Estimated Education Expenditure (R mil / year)	Company	Estimated Education Expenditure (R mil / year)
Anglo American	51.36	Nedbank Group	17.12
Arcelor Mittal	50.73	Telkom	15.04
Truworths	45.6	De Beers	14.72
MTN SA Foundation	35.52	Impala Platinum	13.31
Anglo Platinum	33.40	Old Mutual	12.41
Vodacom	30.55	Transnet	11.25
Standard Bank Group	29.76	Investec	10.68
Sasol	28.8	BAT	9.6
Zenex	28.8	HCI Foundation	9.6
First Rand	28.12	Bidvest	9.28
Liberty Holdings	25.5	Sanlam	9.24
Kumba Iron Ore	25.28	FNB	8.96
Absa	23.29	Sun International	8.96
Pick n Pay	19.2	Woolworths	8.96
Lonmin	18.56		

Source: Trialogue CSI Handbook 2014

Traditionally much of private sector engagement in education has been through provision of bursaries and scholarships to poor learners to attend affiliated private schools, or supplementary enrichment programmes, especially in mathematics, sciences and specialised subjects, offered usually in the afternoons, on week-ends and in school holidays (Besharati, 2015). Support to early childhood education, which is generally under-resourced, is also a popular CSI (see more in section 3.8.2). But, private sector interventions in the education sector are often in the form of infrastructure and facilities upgrading (building new schools and refurbishing old schools), ICT and textbooks provision. In recent years some

companies (ie. Anglo American, SASOL, Rand Merchant Bank, Standard Bank*)* have also ventured in teacher training, both pre-service and in-service (Table 3). The Department of Basic Education (2009), through the 'Adopt a School' initiative has also encouraged private donors to provide additional resources and capacity support to the Dinaledi public schools (see more in chapter 4 and 5). Some companies, such as General Motors, Zenex, Royal Bafokeng, and First Rand have taken their education programming to higher levels of complexity by exploring alternative school development models, governance and accountability systems, teacher development and learner support, and playing an active role in influencing national and provincial education policy. Some of these manifold interventions will be analysed later in this chapter as well as in the empirical work done in chapter 4 and 5. Finally, under the backdrop of a weak public system, the State has often relied on local private companies to be 'contractors' in the provision of educational services such as construction of infrastructure and facilities, and provision of school meals, books, learning resources and uniforms (Pampallis, 2004). Lucrative public tenders have, however, also lead to severe cases of corruption, inefficiency and lack of delivery, as illustrated in the 2012 textbook scandal in Limpopo province (Corruption Watch, 2012; Mail & Guardian 05 July 2012).

Notwithstanding the large volumes of CSI in South Africa's education sector, accountability for such spending remains very poor, due to a compliance-oriented approach to CSI by companies and a lack of interest by senior executives who would rather focus on more critical and pressing divisions of business operations (Besharati, 2015). Overall, monitoring and evaluation of CSI is fairly weak and focuses mostly on inputs and activities (Trialogue, 2011). Very few rigorous longitudinal studies or counterfactual evaluations are conducted on the development outcomes produced by CSI, largely due to the limited technical expertise and paucity of resources invested by the business sector in this field. There is thus little to no evidence on the causal impact and real contribution that the manifold private sector interventions have been making to South Africa's schooling sector. This important knowledge gap will be further explored in the case study in chapter 4 and in the meta-analysis in chapter 5 of this research.

3.7.4 Other partners in education development

In previous sections we have seen how the bulk of financing for South Africa's education sector comes from the state, from parents and from the local private sector. Compared to other African countries, support provided by bilateral and multilateral donors to the South Africa's education sector through Official Development Assistance (ODA) is miniscule (Besharati, 2015), and used more as a strategic extra-budgetary resource to improve systems, share best practices, experiment new models, unlock bottlenecks, add value and play a catalytic role in delivery of services (National Treasury, 2011).

Roughly between 100 and 300 million rand every year (National Treasury, 2009, 2010, 2011) are spent on education projects by a variety of donors such as USAID, DFID, Flanders, Netherlands, France, Ireland, NORAD, SIDA and DANIDA. This is mainly systemic type of support to the National Department assisting in curriculum reform, teacher training, policy development, infrastructure development, administration and institutional capacity development. Other donor support is targeted to specific initiatives such as HIV-AIDs, environmental education, safety and child protection or special needs education (National Treasury, 2013). The European Union's support to the education sector comes in the form of direct budget support to National Department of Education. South-South Cooperation also exists in the sector with countries like Taiwan and Cameroon making some small contributions to South Africa's schooling system. Among the multilateral institutions involved in South Africa's education sector one may find UNICEF, UNESCO, FAO, World Bank, African Union and SADC.

Pre-1994, the non-governmental sector played a very prominent role in the struggle against Apartheid, providing advocacy for policy reform, and reaching out and providing services to the most marginalised communities (Landsberg, Kruger & Nel, 2005). With the ushering of a new democratic dispensation, funding from both international and private donors shifted rapidly from NGOs to the new legitimate government, causing the weakening and closure of many civil society organisations in South Africa (Besharati, 2013b). With still limited capacity in the new public sector (Pampallis, 2003), many of the NGOs changed their advocacy orientation to become more service providers and contractors to the government (Morrow, 2004; Swilling & Russell, 2002). From 1994, the National and Provincial Departments of Education have been outsourcing many education services such as teacher and manager trainings, curriculum development, production and dissemination of LTSMs, developing and implementing assessments, research & evaluation, policy development and school improvement programmes (Morrow, 2004; Pampallis, 2003).

"4.7% (R 916.6 mil in 2015) of national education budget goes to professional services, whereas provinces spend about 8% of their budgets on professional services" (DBE, 2014)

Some of the major education service providers to government, corporate sector and foreign donors include non-for-profit organisations such as Maths Centre, READ Education Trust, Mpower, Bridge, Class Act, Sci-Bono (to name a few) but also private consulting groups such as Khulisa, Eric Schollar & Associates, Genesis Analytics, Education Research Agency, Jennifer Roberts & Associates, all conducting regular research, analysis and evaluations for public or private education interventions. Many think-tanks and research institutes are also present in the sector, conducting analysis and policy advice for South Africa's schooling system. Some of the most prominent of these include the Human Sciences

Research Council (HSRC), JET Educational Services, Centre for Development Enterprise (CDE), Centre for Education Policy Development (CEPD), South African Institute of Distance Education (SAIDE), Catholic Institute of Education (CIE), and the Independent Schools Association of South Africa (ISASA). Many academics, experts and specialised centres at some of the South African universities, such as Stellenbosch, Wits, Pretoria, Tshwane University of Technology, conduct regular research and provide inputs to national and provincial government policy on education. Much of the literature reviewed in this chapter is drawn from the research conducted by many of these institutions and education specialists.

3.8 Interventions to improve learning outcomes

In previous sections of this chapter an overview was presented of the challenges in South Africa's education system. Notwithstanding the great attention given to the sector and the large investments undertaken by both public and private sector, learning outcomes in South Africa still remain extremely poor. There have been nonetheless many programmes, policies and activities undertaken by national and provincial governments, corporate sector, charitable foundations, NGOs and foreign donors, to improve the quality of the country's education system. Naturally, many of the policies and programmes implemented in the field of education may have diverse objectives including expanding access, improving health and nutrition, reducing violence, integration an equity. This analysis, however, will focus on interventions that have as primary objective the improvement of *learning outcomes*.

In sections 3.5 and 3.6, we introduced some of the provincial, national and international learner assessments conducted in South Africa to assess the quality and improvements in the country's schooling system. Often the learner assessments (i.e. SACMEQ, ANA and systemic evaluations) discussed above, have been accompanied by additional questionnaires for learners, educators and school administrators, which allow for the collection of information to explain some of the results and to understand the various other variables which affect learner performance, both internal and external to the school system. Furthermore, other survey data such as the General Household Survey (GHS), the School Management Survey (SMS), the Integrated Quality Management System (IQMS) and the Annual School Survey (ASS) have provided additional data on learners, their communities, their schools (educators, classrooms, infrastructure, facilities, learning and teaching materials). Such statistical information gathered over the years by the Department of Education, Statistics South Africa, the HSRC and other education research institutions, could be analysed in conjunction with the results of learners' test in order to understand better the dynamics at play in the education sector.

Various studies have been conducted world-wide and in South Africa on the factors which impact on learner achievement in public school using some of the above survey data. A number of

econometric models and production functions undertaken in South Africa's education sector (Mouton et al., 2014; Taylor et al., 2013; Simkins, 2010; van der Berg, 2008; Gustafsson, 2007; Prinsloo & Kanjee, 2004) have revealed a number of variables at family, community, classroom, school and system level which have a positive or negative relationship to learner achievement. As an example of just one of these studies, the below table summarises the key results from Stephen Taylor's (2013) school production model on language and mathematics outcomes in primary schools which was conducted with data collected in the National School Effectiveness Study (NSES) (Table 4).

Table 4. *NSES Education Production Function for Learning Gains*

Selected Explanatory Variables	Literacy coefficient	Numeracy coefficient
Student characteristics		
Grade 3 literacy score	0.18***	0.12***
Grade 3 numeracy score	-0.69***	-0.58***
Male	-2.80***	-0,29
Old	-3.77***	-5.21***
Speak English 1-3 times a week	1.55***	2.12***
Speak English 4+ times a week	4.70***	1.86*
English on TV 4+ times a week	3.90***	3.45***
Homework: more 3+ times a week		0.93*
School characteristics		
School SES	1.80***	-2,63
Facilities index	0.24*	

Table 4. *NSES Education Production Function for Learning Gains* (continued)

Selected Explanatory Variables	Literacy coefficient	Numeracy coefficient
Class size	-0.25**	-0.28*
Principal Absent	-1.97*	
Teacher Punctuality	2.52***	2.07*
Timetable available		6.85***
Teacher sick: combine classes		-2.00*
Teacher characteristics		
Teacher Content knowledge		-3.46*
Time spent on assessment: 2-5 hours a week		3.84*
Paragraph writing: none	-3.62***	
Literacy exercises a week: 27+	3.33***	
Short maths exercise a day: 5+ a		5.43*
Maths exercises a week: 76-155		5.15*
Maths exercises a week: 156+		7.93**
Class tests a week: 9+		7.36**
Constant	22.09***	24.61***
R-Squared	0.3336	0.4017
N	8282	8282

*Notes: * p <0.05 , ** p<0.01, *** p<0.0001*

Source : Taylor in Taylor, van der Berg, Mabogoane (2013)

As seen in the table above, some of these school characteristics (e.g., school SES, class size, teacher punctuality, timetable available) have been found to have a strong and significant correlation to language, mathematics and other learning outcomes, and thus results of these studies will be discussed in more detail in the upcoming sections. The above data collected through various school surveys has also provided the basis for the construction of variables used in the regression models and PSM impact evaluation conducted in the next chapter during the Anglo Platinum case study.

The next few pages will review and discuss some of the major types of interventions to address learning shortfalls in public schools, together with the theory of change upon which the programmes and policies have been designed and implemented. Previous research on these types of interventions will be shared in order to understand the dynamics by which the factors, programmes or policies contribute to learner achievement. The categorisation of these education interventions will provide the programmatic typologies and the overall theory of change for South Africa's education sector which will be used in the rest of this research and in the specific evaluation work presented in chapter 4 and 5.

3.8.1 Family, community and society

What emerges consistently from most of the South African education economics literature discussed above (i.e., Mouton et al., 2013; Taylor et al., 2013; Van der Berg, 2008; Gustafsson, 2007) is that social-economic status (SES), family and community dynamics, and other factors outside the school system, are the strongest predictor of learner performance. Systematic Evaluations (2004) show that best performing to least performing schools are strongly associated with the location of the school, whether being in cities, townships, farms, rural or remote areas. Other factors linked to learner achievement include household income, household size, presence of both parents, home language, race, access to television, toys and other learning materials (Gustafsson, 2010; NPC, 2011; Taylor, 2011; Taylor, 2013; Van der Berg, 2008)

In South Africa, like in other developing countries, poverty and inequality are a major determinant of academic success (Fiske & Ladd, 2004; NPC, 2011; Taylor et al., 2003). Educational underachievement, further entrenches the cycle of poverty, which majority of black learners are still trapped in (Spaull, 2013), Many social and economic factors work in tandem to stifle learner development in poor communities. Because of poverty, 25% of children in South Africa (40% in the Eastern Cape province) go hungry as parents cannot buy food, let al.one books, uniform and shoes for their children (OECD, 2008b). Lack of appropriate nutrition affects concentration and cognitive abilities,

which coupled with poor school infrastructure, lack of electricity and running water, do not provide the appropriate environment for effective learning (Bloch, 2009).

The spread of the HIV-AIDS epidemic in South Africa in the early 2000 has had a toll on both learners as well as the educators (Motalaet al., 2014; UNESCO, 2003; Wolhuter, 2006) causing many of them to take frequent leave of absence for health reasons or to take care of ill family members, thus disrupting the basic educational activities (Bennell, 2003; Govender, 2004). Drug abuse, alcoholism, rape and sexual harassment, and violence experienced by learners at home and in the community, have also had a strong impact on the academic potentials of many South African learners (Bloch, 2009). Fractured and dysfunctional families, run by unemployed and illiterate parents or guardians, don't provide the ideal conditions for a healthy and comprehensive development of children in the majority of the poverty-striken communities in South Africa (Luxomo & Motala, 2012). By the age of eight, children from the poorest 80% of households are already far behind the school performance of the richest 20% (Van der Berg, 2008). The disadvantages remain throughout their years of education and stay with them when they enter the labour markets (Taylor et al., 2013).

The National School Effectiveness Study (NSES) showed that household size, availability and frequency of reading books at home does impact on language and mathematics results (Fleisch, 2008; Taylor et al., 2013). A recent impact assessment conducted in South Africa on the Child Support Grant, found that the education of the mother had a significant effect on learning outcomes of children (DSD, 2012). A well-educated parent closely engaged with the education of their children does show to have an impact on both school enrolment and learning outcomes (Epstein, 2001; Fleisch, 2008; Soudien, 2007). PIRLS data (2006) highlighted that regular parental engagement in school activities – through volunteering, sporting and social activities, parent-teacher meetings – does influence on academic results of learners.

Active and concerned parents also assist in improving accountability and functionality of public schools. As discussed earlier, SGBs are the main channel where parents can directly engage and influence in the operation of the schools. Experience has shown, however, that SGBs are more functional in more affluent and previously white schools where community capacity is stronger, rather than poor areas where parents have less time and resources to dedicate to school improvement and where teacher unions end up playing a more prominent role (Taylor, 2011).

Both government and non-governmental partners have implemented a range of programmes aimed at strengthening school governing boards (SGBs) and local stakeholder engagement in monitoring the quality of education services in South Africa. This acknowledges the important role parents and community play in the learning process of their children both inside and outside the formal school space.

The introduction of report cards for parents and school managers through the ANA process has also been a critical tool to improve accountability of teaching and learning processes in the last five years.

The Joint Education Trust (JET) since the early 2000s has been piloting, in its various school development programmes various strategies for parental and community engagement. Aside from information tools such as report cards and regular dashboards on school inputs, outputs and processes, JET has also encouraged the formation of study groups led by volunteer parents who provide additional tutoring to children in the communities. All of these initiatives have had a marked effect on improving learner performance (JET, 2008).

3.8.2 Early childhood development

Closely linked to parent's education and social-economic conditions (discussed in the section above) is the role of early childhood development (ECD). Most of a child's deductive ability in-fact is developed before he/she starts formal schooling (DPME, 2012; van der Berg et al., 2013) ECD can contribute substantially to children's cognitive, emotional, social and physical development (Jujurha Education Centre, 2010) and assists them greatly in their academic performance throughout their schooling years. ECD has a lasting impact on the overall population and it is a fairly low cost intervention to raise education standards (Barnett & Steven, 1995).

During apartheid ECD was almost unheard of in black communities (NPC, 2011). African students would enter school unprepared and disadvantaged compared to similar peers. This would have a long-term detrimental effect on their entire education, causing difficulties to adjust at school, resulting often in poor performance and early drop-out. Historically, ECD was undervalued and received less attention by government than higher education for example. Prior to 2009, less than 16% of children of South Africa were enrolled in ECD programmes. The few ECD facilities in poor and rural areas were characterised by poor infrastructure, the lack of learning materials, toys, food, and most teachers were not adequately trained (only 12%) (UNICEF, 2011).

Because of the lack of attention by government, international donors, NGOs, community-based organisations and private companies have been supplementing the funding gap to this sector. Early childhood development (ECD) has in fact constituted 20% of the CSI in education (Trialogue, 2011), and some of the major exponents of such programmes include ABSA, First Rand, BHP Billiton, SASOL, Murray & Roberts, Nedbank and Rio Tinto, to name a few.

With the Delivery Agreements signed in 2009 by South African President and the Minister of Basic Education, the Department of Health and Social Development, Women, Children and Disabilities, ECD has achieved a new political imperative. Increase in the coverage and quality of ECD is Output n. 3

of the Basic Education Delivery Agreement and per capita spending on pre-school education has increased to R3100 in 2009/10 financial year (DBE, 2011). The Department of Basic Education has renewed its commitment to engaging more South African learners in a Reception Year (Grade R) in most public schools and has committed in its Strategic Plan to achieve 100% enrolment in Grade R by 2014. By 2013, 95% of 6-year olds across the country were enrolled in grade R and 75% of the children 5 years old and younger old were participating in some kind of ECD (Taylor et al., 2013)

Recent evaluations around ECD have included a public expenditure tracking survey (PETS) conducted by UNICEF (2011) in the sector, as well as study commissioned by the Department of Basic Education and the Department for Performance Monitoring and Evaluation (DPME) to assess the impact of the recent grade R campaign (Van der Berg et al., 2013). The Stellenbosch grade R study utilised a dataset that combined ANA results for grade 2, with Education Management and Information System (EMIS) and SNAP datasets on learners and schools, into a fixed effect regression model. The report (Van der Berg et al., 2013) found that pre-schooling did have an effect on learning, especially language results, but this varied according to the school quintile, provincial location and the abilities of the individual learners. The study confirmed that impact of grade R is dependent on the quality of the curriculum, teachers, resources, and parental support.

3.8.3 Language Policy

According to an HSRC study, after poverty the biggest factor influencing school results in South Africa is language (Prinsloo & Taylor, 2005). Taylor et al. (2003) also found that practice of speaking and exposure to reading at early ages had a large impact on learning. Without a good command of language one struggles to grasp more complex mathematical and scientific concepts (Botes & Mji, 2010). In Taylor's (2013) school production model exposure to English (i.e. speaking frequently at home or watching television) had an impact not only on literacy scores but also on mathematics results. Studies on school development in South Africa show that there is a strong correlation between literacy test scores and numeracy test scores (Simkins & Perreira, 2010). In another CDE study (Simkins, 2010), multivariate regression on mathematics scores show that after race, the biggest factor influencing results is home language (with an impact factor of 11%). Previous literature (Myburg, Poggenpoel & Van Rensburg, 2004) has highlighted that using mother tongue as the Language of Learning and Teaching (LoLT) in schools will assist learners in their academic achievement.

South Africa, however, presents the challenge of having numerous cultural and language groups in its territory and the language used at school is a highly political issue. The constitution of South Africa (1996) has decreed 11 official languages and has given the right for any learner (and their parents) to

choose one of the 11 language as their language of instruction. Within its ability, the state has the duty to provide teachers and study material in any of the official languages, especially if there are 40 or more learners in primary school that demand such. The South African Schools Act (1996) takes a communitarian approach (Woolman & Fleisch, 2008) and places the power on the local SGB to choose the Language of Learning and Teaching (LoLT).

Post-1994, the educational policies have tried to gradually address the Apartheid inequalities in language of instruction, however still today LoLT in South African schools remains a highly complex matter. To illustrate better the point, 65% of schools teach in English, 12% in Afrikaans, 7% in isiZulu and 5.5% in isiXhosa while 25% of South African learners speak at home isiZulu, 20,5% isiXhosa, 10% Afrikaans, and 7% English (DBE, 2010).

The Language in Education Policy (LiEP, SASA, 1996) emphasises choice by learners, parents and schools in selecting the main LoLT as well as an additional second language to promote communication and intercultural bridging, critical to South Africa's nation-building processes. Underpinned by notions of equity, redress but also practicality, LiEP and the National Curriculum Statement (NCS, 2008) encourages the use of home language especially during the foundation years (grade 1 to 3), while strengthening at the same time a second additional language (typically English or Afrikaans), which learners switch to in grade 4 (Standen & Boskor, 2016).

The language policy in South African schools has generated much debate between those who promote the use of English from the beginning of schooling and those who advocate the use of African languages to promote redresses empowerment and contribution by Africa to knowledge and the sciences (Alidou et al., 2006; Taylor, Gamble, Spies, & Garisch, 2012). Although the general consensus is to encourage the use of indigenous languages in South Africa's schooling system, there are a number of practical constraint that still remain, such as the lack of books and materials in African languages (which often do not possess the appropriate academic and technical terms), the limited number of trained teachers who teach in mother tongue, and even a certain degree of resistance by local communities who undervalue the use African languages in the formal schooling system (Taylor et al., 2013; Van Staden & Bosker, 2014).

As a result black learners are not properly prepared for the sudden language switch in the transition from foundation to intermediary phase (grade 4), and thus 81% of South African learners, who speak at home one of the 9 African languages, are forced to suddenly switch to English or Afrikaans as their medium of instruction; thus having a deleterious impact on learning outcomes (Taylor et al., 2012). Van Staden and Bosker (2014) have shown that learners writing a grade 4 language assessment in home language perform 29 points (3/4 of a year difference) better than those who write the test in English or Afrikaans. This echoes some of the earlier studies (Howie et al., 2007; Spaull, 2011) that show that

English and Afrikaans speaking learners perform better in literacy than their peers with other home languages; but this might also be masked by other SES factors (see previous section) and educational constraints in black schools (i.e. poorly resourced and badly trained teachers) (Vorster, Mayet & Taylor, 2013).

In recent years, the DBE has revised the curriculum to reduce the subjects in Foundation year to 6 (instead of 8) and include a stronger element of English/Afrikaans as a second language for all learners. Generally speaking, more rigorous studies are required to shed more knowledge in this critical arena of education policy. An impact evaluation of the LiEP (1996) would be interesting to do, however very challenging due to its systemic nature. Lighter impact assessments on literacy programmes have been conducted by various organisations and evaluators (see more in chapter 5), including a study by Taylor (2011) which demonstrates that children who spend more time doing reading and writing exercises, perform better at school.

3.8.4 Learning and teaching support material (LTSM)

One of the most straightforward ways to improve the quality of education is the provision of materials, technology and other resources to schools to enhance the teaching and learning experience (OECD, 2008b). Examples of such can be textbooks, workbooks, study guides, lesson plans, science-kits, computers, multi-media accessories, educational toys, reading books, etc. Crouch and Mbogoane (2001) have insisted on the importance of such 'cognitive resources' for effective learning to occur. However, Taylor (2013) and Van der Berg (2008) have highlighted that provision of LTSMs have an impact only with strong educators and school leadership.

The most classical Learning and Teaching Support Materials (LTSMs) have been textbooks. In line with the National Curriculum Statement various textbooks have been approved and distributed in all languages to all schools in South Africa. Particularly noteworthy has been the contribution of the Shuttleworth Foundation in the development and production of a large series of specialised textbooks that have been distributed throughout the country (Narsee, 27 October 2012, personal interview).

Following the failure of the OBE experiment (Motala et al., 2014), discussed in earlier sections, and the introduction of Curriculum Assessment Policy Statement (CAPS, 2011), educational approaches in South Africa have gradually moved to more prescriptive process-oriented models of school development (Fleisch, 2002). There has thus been also a move towards a more prescriptive curriculum, leaning on the use of scripted and standardised lesson plans (de Clercq & Shalem, 2014). This has given rise to the 'workbooks' approach in South Africa's schooling system (de Clercq & Shalem, 2014), with some notable experiments and pilot projects undertaken in Gauteng (i.e., GPLMS, READ, B2B) and in

Limpopo province (i.e., PMRP), which will all be discussed in more detail in the meta-analysis in chapter 5. The rationale of using workbooks rather than textbooks is that they provide pacing and guidance for both teachers and learners, by providing weekly lesson plans and tailored-made exercises. Challenges of language difference and multi-grade classes would also be alleviated as each learner would have their own workbook based on his/her specific needs (Schollar, 2015).

In 2010 the President announced the government's Workbook Project, an initiative implemented by the Department of Education that would provide all learners in South Africa user-friendly standardised workbooks in all official languages. The distribution of the first batch of workbooks started in 2011 with the provision of 24 million mathematics and literacy workbooks for grade 1 to 6 learners to the poorest quintile schools (Presidency, 2012). Since then every learner from grade 1 to 9 has been receiving 2 high-quality mathematics and language workbooks, which have provided a weekly structure to curriculum delivery and facilitated teachers in their lesson planning (Spaull, 2015).

A common problem with textbooks, workbooks and other LTSM provided through the public sector is the inefficiency in the distribution mechanisms, the late arrival of materials to the schools and the corruption which often occurs in such processes. Recorded breakdowns in the LTSM provision in the Eastern Cape and in Limpopo (Chisholm, 2013; Veriava, 2013) are a testimony of such systemic problems. DBE and UNICEF have commissioned the Australian Centre for Education Research (ACER) to conduct a summative implementation evaluation of both the textbook and the workbook distribution (ACER, 2013). On a more systematic level, the School Monitoring Survey (SMS) regularly checks if schools around the country receive their assigned LTSM on time.

Aside from the textbooks and workbooks, the simple provision of reading books for school and community libraries is an important contribution to improved literacy. Only 8 % of public schools in South Africa (mostly former model C schools) have functioning libraries (DBSA 2008). National and International studies (Lee et al., 2005; Bhorat & Oosthuizen, 2008; Taylor, 2007; READ, 2010) shown well staffed and well funded libraries adds 8 to 18 percent points to average learning outcomes of schools. The provision of libraries can thus be a cost-effective way to improve learning results in South Africa's education system (Equal Education, 2011).

The private sector is also heavily engaged in providing different forms of LTSM to the schools of South Africa, supplementing the resources provided by the state. A prominent corporate campaign in South Africa is the *Rally to Read* led by the McCarthy Group, where trucks and 4x4 vehicles travel every year to remote parts of the country to distribute books and other material to schools. Other popular forms of provision of LTSM by the private sector include the provision of science labs to schools (Besharati, 2015). Some of the more active organisations in support to science education in schools include the Gauteng Education Development Trust, Sci-Bono Center, Anglo American, Telkom and various oil

companies (Sasol, Caltex, BP, etc.). Often publishers and print media provide grade 12 learners with LTSM in preparation for the NSC such as example exam papers and learner study material the likes of Math 911 produced by Liberty Life, Power Your Future supplement of the Sowetan, Read Right Edition of the Sunday Times and Matric Matters of Independent Newspapers (Besharati, 2015).

What is increasingly becoming popular in South African schools is the use of ICT facilities, computer software and multimedia material to enhance the learning experience. Examples include companies such as Cell C developing mathematics revision games for mobile phones, or the Gauteng Education Development Trust piloting educational software such as iPower2 and Learn-things. On the media front several tutorial programmes have been developed such as the popular Mindset program on DSTV, Study Mate on SABC and educational films screenings at Sterkinekor cinema. From the side of government, the Teachers Laptop Initiative, managed by the PED and the ELRC, has provided educators with financial support to purchase computers, teaching and administration software, internet connectivity, thus enhancing their teaching capabilities (Gauteng Education Development Trust, 2013) Much of these ICT interventions are promising and still in pilot stage, thus experimental evaluations would also be appropriate measurement instruments to test such initiatives. Some examples of RCTs on technology and computer-aid learning (CAL) (Bohmer, 2014; Botes & Mji, 2010; Louw et al., 2008) are discussed in more detail in the meta-analysis in chapter 5.

3.8.5 The central role of teachers

The National School Effectiveness Study conducted in 268 schools which followed 16,000 children from grade 3 (2007) to grade 5 (2009) showed that, after home characteristics, teachers have the strongest influence on learning outcomes in South Africa (Taylor, 2011). The success of the education system lies mainly in the critical interaction that occurs between teachers and learners in the classroom (World Bank, 2011). In his school productivity OLS model, Gustafsson (2007) highlights the influence of teachers and managers time, teaching methods, textbooks, nutrition and infrastructure on learning outcomes. Simkins (2010) emphasises that the low output of school graduates of mathematics and science is due to poor quality teachers and ineffective schools. And more, Taylor (2011) has demonstrated the positive correlation of teacher knowledge, curriculum coverage, assessment planning on literacy and numeracy results, and has stressed how simple improvements in the quality of teaching are the most cost-effective way to improve South Africa's education system.

The political arena has also echoed the concerns from academia. President Zuma in his third State of the Nation Address (SONA, 2011) has emphasised:

"We reiterate our call that teachers must be at school, in class, on time, teaching for at least seven hours a day. The administration must ensure that every child has a textbook on time, and that we assist our teachers to create the right working environment for quality teaching to take place."

The issue of teachers and relationship to quality education is fairly complex and is related to issues of quantity, knowledge and accountability in the educator profession. Thus, each of these dimensions will be discussed in a bit more detail.

a) Teacher quantity

In the late 1990s the government went through a process of rationalisation and reassignment of teachers among the various schools (urban-rural, rich and poor) to attempt to achieve more equitable distribution of educators and learners across the public school system in South Africa (Fiske & Ladd, 2004). From pre-1994 black school classroom sizes of over 50 students (Woolman & Fleisch, 2006), South Africa's public schools have stabilised at an average educator-learner ratio of 1 to 29.4 (DBE, 2013). However, many differences still remain between high quintile and the low quintile schools in the quantity and quality of teachers (World Bank, 2011). As a result of the SASA (1996) policy frameworks, the wealthier schools, are able to raise additional funding to recruit more educators to spread better the teaching workload. Some government interventions, like the Dinaledi, and initiatives by private sector and NGOs have endeavoured to support selected schools by providing additional teachers or teaching assistants for short or long periods in order to strengthen the teaching of specific subjects (Blum et al., 2010; Besharati, 2014).

While post-1994 there was an over-supply of black teachers for historical reasons (discussed later), over the years many of the educators from the previous generation retired, moved to other professions, or dropped out of the system for disciplinary and health reasons (Wolhuter, 2006). At the current pace of growth, there are not enough teachers (especially of mathematics and sciences) present in South Africa to meet the demand of its economy (CDE, 2011, SACE, 2010). Every year, 25,000 new teachers are required and the country is able to produce only 10,000 (Deacon & Simkins, 2011). Crouch (2002) had already foreseen that in South Africa by 2015 there was going to be a shortage of between 11,000 and 57,000 teachers. Of teachers who are currently in the system, many are also being improperly

utilised. In the Eastern Cape, of the 16,500 trained mathematics teachers only 7000 are teaching their own subjects (Deacon & Simkins, 2011).

Post-1994, many of the best teachers have migrated to other countries or other professions with better pay or with better public prestige. The educator corps is highly feminised - 80% of primary school and 51% of secondary school teachers are women (Crouch, 2002; UNESCO, 2006). The teaching profession is unfortunately highly undervalued by black South Africans (Habib, 2013) and does not attract young capable and ambitious individuals (Cosser, 2009). Principals complain that it is difficult to attract good individuals to the teaching profession (Christie, Butler, & Potterton, 2007). Some have suggested that immigration policies should be revisited to encourage the importing of qualified teachers from other African and developing countries like Uganda, Zimbabwe and India (CDE, 2007; CDE, 2010a), which is already is the case of many of the mathematics and science educators in South Africa's secondary school system (Besharati, 2014; Spaull, 2009, 2013). The issue of better performing foreign mathematics and science teachers has also been touched upon in the empirical work conducted in chapter 4 in the case study in Limpopo and North West province.

One of the Department of Education's prime initiatives to increase the number of new educators, particularly in rural and poor areas, has been the 'Funza Lushaka' (Teach the Nation) scholarship, which was instituted in 2007. This bursary scheme covers tuition, accommodation, meals, books, materials and living allowance of black students who decide to take on the teaching profession, particularly in priority subjects (language, mathematics and science). On completion of their training, bursars are required to work for the provincial education departments and be deployed in needy areas for the same amount of years they received financial support. Since its inception Funza Lushaka has provided more than R 1.1 billion to more than 30,000 students resulting in 6,000 new graduates by 2012 (DBE, 2012). The Department of Basic and Higher Education and Training in collaboration with the Department of Planning, Monitoring and Evaluation in the Presidency are currently undertaking a national evaluation of this important bursary scheme aimed at raising more teachers in South Africa's schooling sector.

a) Teacher quality and training

Having a sufficient number of teachers in the system is not enough to resolve the problems in South Africa's education sector. The quality of teachers is an even more important factor in the learning process (Hopkins, Ainscow & West, 1994; Hanushek & Woessmann, 2012; Scheerens, 2000). "Any education system stands or falls by the quality of its teaching profession" (Wolhuter, 2006, p. 124). The quality of an education system can never exceed the quality of its teachers (McKinsey & Co, 2010).

Under the apartheid system, the teaching profession was one of the few jobs available to the black middle class, and a vehicle for upward social mobility (Wolhuter, 2006). While there was an over-supply of black teachers during Apartheid, the legacy of the Bantu Education Act (1953) ensured that black educators did not receive the same quality of training and support as their white counterparts (Bloch, 2009; Wolhuter, 2006). Furthermore, in a fragmented education system run by different Bantustan states, the diverse cohorts of teachers were trained under different systems by a variety of provincial institutions (Chisholm, 2009; Sayed, 2004). The rough division of responsibilities for teacher training was that universities would train secondary school teachers, while colleges would train primary school educator, creating thus a '1st and 2nd class' tier of educators in the system (Chisholm, 2009). Like for the rest of the social system under the fragmented apartheid regime, there was a large variety of competencies and qualifications among educators in South Africa, divided mostly on racial lines (2006).

Table 5. *Teacher Qualifications in South Africa in 1988*

Qualification	% Black	% Coloured	% Indian	% White
University degree	5	13	43	32
Secondary school	62	63	57	68
Below secondary school	32	24	0	0

Source: Wolhuter 2006, Teacher Training in South Africa

During the political transition period in the 1990s, there was a gradual process of rationalisation and restructuring of the former provincial teacher training colleges which was completed by 1997 (Chisholm, 2004; Sayed, 2004). Efforts were made to shift all teacher training to the universities of the country. The high quality colleges were incorporated/integrated as rural extensions of the universities, while the poorly performing teacher training technicons were converted into FET schools or provincial government training centres (Chisholm, 2009). Some 150 teacher training colleges active in 1994, had reduced to 50 by 2000, and by 2006 all teacher training was occurring almost exclusively via the 24 South African universities (Wolhuter, 2006, Table 5).

Over the years there has been much debate in the academic and policy space about re-opening the old teacher training colleges, particularly by the unions and educators nostalgic of the social bonds that the colleges would provide during the pre-1994 era (Chisholm, 2009). Some have also argued (Chisholm, 2009; Golden & Daniel, 2007) that universities are too expensive, cater for the elite, and impart

theoretical knowledge that is not always attuned to the practical reality of primary schools. Pelzer et al. (2005) has also argued that re-opening the former colleges can address some of the current shortage of teachers, particularly in the foundation phase, in mother tongue languages and in mathematics and sciences. Some of these issues have also emerged from principals and teacher during the field work in Limpopo and North-West discussed later in chapter 4.

The pre-1994 legacy has thus left the South African schooling system with a vast majority of old generation teachers (above 40 years old), who have been (poorly) trained under the previous apartheid regime (CDE, 2011). As a result, the majority of South African teachers have poor subject knowledge but also limited pedagogical skills for effective teaching (Taylor, 2009; Spaull, 2011). Various research (Carnoy et al., 2011; Spaull, 2011; Taylor, 2009) have highlighted that South African teachers do not know the subjects they are teaching. In a JET study, teachers of grade 4 and 5 were given a mathematics test for grade 6 learners and only 12% of the teachers could answer correctly to all five basic questions (Taylor, 2011).

The National School Effectiveness Study showed that low performance of South African learners was strongly associated with poor content knowledge by their teachers (Taylor S in Taylor et al., 2013). It has been estimated that mathematics results could increase 35 to 45 per cent if simply teacher knowledge could be raised in an effective manner (Taylor, 2011). Thus most common interventions implemented in South Africa to improve the quality of teaching have been various forms of pre-service (PRESET) and in-service (INSET) educator training programmes.

A number of studies in South Africa (Crouch & Mabogoane, 2001; Vinjevold & Crouch, 2001), have shown that learner performance is closely correlated to teacher qualification. Taylor and Vinjevold (1999) have also discussed also how poor subject knowledge and poor pedagogical skills stems from unsatisfactory training and qualifications (See also JET, 2005). In 2006, approximately 80,000 teachers did not have an appropriate professional qualification (Wolhuter, 2006). In 1990 only 53% of educators were deemed qualified, but in 2009 this number rose to 94% (Woolman & Fleisch, 2009).

The requirement to be registered as a professional educator in the South African Council for Educators is the possession of NSC (matric) plus at least three years of relevant teacher training. The NEPA (1996) provides the Department of Education, to develop the national policy on teacher training, including the relevant curriculum framework. Norms and standards for teacher training were established in 1998 and later revised also in 2000 (Brunton, 2006). Most pre-service trainings in South Africa occurrs through the Postgraduate Certificate in Education (PGCE) programmes at various universities and colleges.

The quality of pre-service educator development programmes has been, however, highly debated, as only one third of training institutions qualify for accreditation (CDE, 2011). In a report on the various

educator programmes offered in South Africa, the Council on Higher Education (CHE, 2010) concluded that billions of rands spent on teacher training did not result in improved quality of teaching, and many educators still lack substantive knowledge in specialised subjects.

The weaknesses and challenges of the current pre-service training schemes was also acknowledged during the Teacher Development Summit in 2009, which brought together South African government departments, SACE, teacher unions and other education stakeholders (Sayed & Motala, 2012). Following the Summit, a new 'Integrated Strategic Planning Framework for Teacher Education and Development 2011-2015', was adopted which outlined a national plan for recruitment and preparation of teachers, induction in the workplace and continuing development. The Plan proposed the establishment of the National Institute for Curriculum and Professional Development (NICPD) and various Provincial Teacher Development Institutes (PTDI) that would jointly work together in the certification of new teachers, upgrading of current teachers, introduction of curriculum changes and continuation of life-long teacher development.

Parker (2006) has argued that above the reform of pre-service training, re-skilling of South African educators through in-service training is of highest priority. Wolhuter (2006) has further reiterated that more than 300,000 teachers in South Africa need to be re-trained. Several different INSET interventions are regularly implemented by both National and Provincial Departments of Education. These usually consist of upgrading of qualifications, knowledge or skill-sets of teachers through different models of training implemented over week-ends, evenings, school holidays or several month of block-release at colleges and universities. The main teacher upgrading and re-skilling programme has been the Advanced Certificate in Education (ACE), however this has also been gradually phased out.

Effective teacher development programmes often also have a component of classroom visits by subject and pedagogical specialists, teacher coaching, monitoring, and regular sharing of experience between professionals (see some examples in chapter 5 meta-analysis). The corporate sector has also been offering its own teacher development programmes, through initiatives of Anglo American, Nedbank, SASOL, Rand Merchant Bank, Standard Bank, First Rand, Transnet, and Shuttleworth, to name a few.

Ono and Ferreira (2010) conducted a study on the various in-service teacher training (INSET) programmes which are available in South Africa in different shapes and forms (courses, conferences, seminars, workshops) The report concludes that most INSET have been ineffective, too brief, incoherent and de-contextualised from real classroom situations. Nick Taylor (2011) has also voiced strong views about the ineffectiveness of most INSET programmes that run on week-ends, holidays or afternoons. He has advocated for residential block-release programmes (of at least one week) in which teachers are immersed in intense training, while their classes are covered by substitute teachers. This is the INSET

model used by the Cape Teaching and Leadership Institute (CTLI), MASTEC college in Limpopo and the Sci-bono Centre in Gauteng (Taylor, 2011; Taylor et al., 2013). Chapter 4 and chapter 5 will discuss in more detail some of the impact evaluations conducted on some of the above mentioned INSET teacher training programmes.

b) Teacher accountability and unions

Even with enough teachers with the right qualification placed in all schools, one of the major challenges of the South African public education system is poor time management practices and lack of professionalism among educators. While there are a lot of capable, committed and caring teachers in the systems, there are also a lot who simply do not do their job, as they are protected by their unions and function within an accountability vacuum (Bloch, 2009). The South African education system is very week on accountability mechanisms, which are necessary so that educators are incentivized to perform their work properly and to constantly improve their teaching abilities.

The struggle against the apartheid authorities of the time, had left a profound mark on discipline of both learners and teacher in black schools, which was difficult to rectify in the post-1994 era (Fleisch, 2002). This 'laissez-faire' culture in South Africa's education system was summarised by Jansen (in Sayed et al., 2013) as the 'apartheid legacy of disgust, which created the post-apartheid inheritance of distrust'. This among other things, has affected the current 'time on task' of bother teachers and principles (Taylor, 2011), which in turn has significant impact on school productivity in South Africa.

The National School Effectiveness Survey (Taylor et al., 2013), discussed earlier, has highlighted that teacher commitment (through punctuality and presence), lesson planning, effective curriculum coverage, and regular homework and assessments, have a clear (and obvious) correlation to learner achievements (Taylor, S, 2013). A challenge in South Africa however has been getting teachers to school, to class and to cover the curriculum (Taylor et al., 2013). A comparative study between similar schools in South East Botswana and North West of South Africa (Carnoy, Chisholm & Chilisa, 2012), showed that Botswanan learners performed better than South African learners, partially because Botswanan teachers had covered 60% of the curriculum, while their South African counterparts had covered only 40%. This points to systemic problems with time management within the South African educator corps.

The SACMEQ 2 data also point to very high levels of teacher absenteeism and late arrival, especially in poorest quintile schools in South Africa (Van der Berg & Louw, 2007). Several other South African studies (Reddy et al., 2010; Langa & Du Toit, 2007; Chisholm et al., 2005) on teacher absenteeism showed that teachers are found teaching on average 3,5 hours a day instead of the required 6,5 hours. Teachers come late to classes, leave school early, skip classes (SACMEQ, 2000) and are often

on leave on Mondays and Friday (20% of teachers in the study). These patterns worsen towards the end of the month. Reddy et al. (2010) have also discussed the abuse of 'sick leave' by teaching and non-teaching personnel in South African schools. Many school days are also lost because of periodic strikes and trade union meetings held during class hours. Due to these factors, but also to unproductive practices such as copying notes on blackboard (because of shortage of textbooks), cumbersome administrative duties, teachers end up teaching only half of the time they are expected (Chisholm, et al., 2005). All these practices in the long run cause learners in African schools to effectively lose 3 years of schooling in comparison to their peers in affluent schools, thus having obvious consequences on the learning results and future employability (Chisholm et al., 2005; Langa & Du Toit, 2007; Reddy et al., 2010).

Teacher unions periodically make the argument that to raise the quality of education in South Africa, educator salary must be increased. However, studies commissioned by OECD (2008b) and by the National Planning Commission (2011) have indicated that, compared to other professions requiring similar qualifications and work-time, South African public teachers are very well paid and have better working conditions (see also Crouch, 2002). In fact, the high salaries could sometimes even attract the wrong people to the profession for the wrong reasons. Nonetheless, appropriate incentives systems need to be created to motivate educators to stay in the profession, perform better and improve learning outcomes. A study conducted by UNICEF (2009) concluded that parents, teachers and principals agree that educator pay should be linked to performance, however political opposition (particularly by unions) often arise when trying to implement such schemes.

Teacher unions has emerged to be a major stumbling block to the much needed reforms in South Africa's education sector (Bloch, 2009). Historically, trade unions in South Africa have been very powerful and at the heart of the political and economic emancipation of black communities. The South African Democratic Teachers Union (SADTU), which represents over 250,000 teachers (SADTU), has been very militant in the 1990s defiance movement, pressurising principals and DoE officials to resign from their posts (Fleisch, 2002). Other influential teacher unions include National Professional Teacher's Organisation of South Africa (NAPTOSA) and *South African Onderwyser Unie* (SAOU), with 95,000 and 41,000 members respectively (Govender, 2004). The umbrella organisation, the Congress of South African Trade Unions (COSATU) is a central party to the ANC-led tripartite agreement that forms the post-1994 ruling government. Hence, SADTU and other teacher unions are very closely embedded with the Education Labor Relations Council (ELRC), the South African Council for Educators (SACE) and various other policy-making forums of the Department of Education (Govender, 2004), to a point that is no longer healthy for accountability and reform of South Africa's education sector (Spaull, 2015).

Unions have strong influence on school hiring, promotion and disciplinary measures of teachers, particularly in poor communities where parents are less involved in school governance (Taylor et al.,

2012). Unions exert power on deciding appointments and maintaining in position teaching and non-teaching staff. Such practices undermine the merit-based system, as ill-equipped people are often appointed to important jobs, thus affecting poor quality of services. Impunity is also endemic in the public school system as disciplinary procedures for dismissal of non-performing teachers are complex and time-consuming, thus rarely pursued (NPC, 2011). Such environment sends out wrong messages - that knowledge is less important for professional development, than for instance the nourishing of political relationships (Taylor, 2011). All of this promotes a predatory culture of nepotism and patronage (Spaull, 2015; Taylor et al., 2012), which is detrimental to much needed improvements in the education sector. Taylor (2011) argues that for a major change to occur in the South Africa's education system, engrained nepotism needs to be addressed from the highest to the lowest political structures.

To revive the South African schooling sector there is an urgent need to put into place strong accountability systems, that have the right balance of rewards and sanctions (or carrots and sticks) in order to work. As Nic Spaull so bluntly put it:

"Accountability is another word for consequences. When there are no consequences for non-performance there is no accountability. It isn't complicated. Currently there are no consequence for non-performance. Not for teachers, not for principals, not for district officials, not for union leaders, not for bureaucrats, not for DDGs." (Spaull, 2015)

Many previous attempts to reinstate accountability mechanisms in South Africa's schooling sector, such as the Development Appraisal System (DAS), have been largely unsuccessful. In 2003, after negotiations between the Education Labor Relations Council, the Department of Education and teacher unions, the Integrated Quality Management System (IQMS) was instituted, which combined school assessment, through the Whole School Education Policy, with educator assessment, through the Development Appraisal and the Performance Measurement System (de Clercq, 2008). The IQMS provided a 360-degree assessment of teachers via a) self-reflection, b) development support group (peers), c) school managers, and d) district moderators/inspectors. It performed both an internal formative role (learning, exchange and personal development) with an external summative role (promotions, awards, and salary increases) (de Clercq, 2008). The IQMS aimed to improve accountability and professionalism of teachers through periodic spot visits of District monitors (7000 in 2011), provision of yearly awards and salary notches for best performing teachers.

Like its predecessors, the IQMS did not work and was subject to much criticism (de Clercq, 2008). As expected teachers and unions opposed the system, calling it unfair and inappropriate (SADTU, 2005). It was in fact based on many assumptions that didn't match the reality of the South African school

system: teachers did not have the work ethos and the collegiality (Wadvalla, 2005) required to improve teaching practices; and heads of department, principals and district officials didn't have the capacity to monitor and provide adequate support to the teachers (Marneweck, 2007; de Clercq, 2008). It was a very subjective exercise as different stakeholders were driven by different incentives and biases (de Clercq, 2008). Rather than leading to improvement, it became yet just another administrative and bureaucratic routine (Ryan, 2007).

In well-performing schools IQMS was to some extent working, but in dysfunctional schools, IQMS became a cumbersome, time-consuming and fruitless exercise, which didn't bring any benefit and therefor not treated seriously (de Clercq, 2008). In an official review of the policy, Class Act concluded that the IQMS is an unreliable and invalid process through which educators are assessed and rated irrespective of their learners' achievement (Marneweck, 2007). In fact every year all teachers receive a 1% salary increase regardless of their professional performance (Taylor S, 2012; Wills, 2015). Nick Taylor (2011) has reiterated that the IQMS and any other incentives and accountability system will continue to fail for the lack of political will and implementation capacity by education managers.

A less threatening instrument that is currently practiced by government authorities but also by the private sector (ie. Massmart, Dell, Optima & Prima, Intel, Investec, Telkom, Old Mutual, Vodacom) are teacher awards for good performance and good results. This form of incentive though small can also be more easily assessed through standard impact evaluation techniques (discussed in the previous chapter).

In conclusion, it needs to also be acknowledged that educators cannot be held accountable if they do not have capacity and they do not receive enough support (de Clercq, 2008; OECD, 2008b; Spaull, 2015). One initiative to support teachers in their functions was the Foundation for Learning Programme (FLP) run by the Department of Education between 2007-2011 in the GET band. The FLP provided schools and teachers with clear directives with regard to optimum use of time, daily teaching activities, basic recommended resources and assessment packs. This was coupled with increased monitoring, evaluation and support by provincial and district education offices. In 2011 the Minister of Basic Education Angie Motshekga launched the National Education Evaluation Development Unit (NEEDU) to research strategies to improve performance of teachers and the quality of education in South Africa, and in 2012, she appointed Dr. Nick Taylor, former CEO of JET and one of the strongest critics of SA's education system, to head up the new unit.

3.8.6 School management and principals

Very closely associated to teacher professionalism is the role of school principals (Taylor et al., 2003). Research (Simpkins, 2010; Taylor, 2011) has shown that one of prime strategies to achieve improved

school results is increasing internal accountability of teachers and management. After discussing all the various factors which affect the quality of education in South Africa the National Planning Commission concluded in its Diagnostic Report (2011, p.15):

"…without dismissing any of these factors, our conclusion is that the main problems lie in the teacher performance and the quality of school leadership."

Bush (2007) has highlighted that the central role of the school principal is to promote an environment where effective teaching and learning occurs. He further elaborates the functions of an 'instructional leader' (Bush, Joubert, Kiggundu & van Rooyen, 2010) that is to oversee curriculum implementation, ensure availability of learning material and that HoDs monitor the work of educators and their lesson planning. Model principals promote high standards of teaching, conduct classroom visits for quality control, and regularly monitor overall learner performance in the school (Bush et al., 2010).

A number of studies conducted in South Africa (Bush & Heystek, 2006; Carnoy et al., 2011), have found however that school principals do not understand their role as educational leaders. Instead of focusing on supporting and incentivising teachers and monitoring curriculum coverage, they spend most of their time on HR/financial management, policy issues, disciplinary matters, administrative chores and reporting to the Department of Education (Bush, 2007; Chisholm et al., 2005). Committed school managers, on the other hand, are expected to keep teachers accountable, mentor less-experienced staff, inspire learners, ensure books and material are delivered on time, use efficiently the meagre resources and facilities at their disposal, run disciplined schools, resolve conflicts, involve parents in school affairs and engage other community stakeholders (Taylor et al., 2012). Leithwood et al (2004, p.7) have further stressed:

"There are virtually no documented instances of troubled schools being turned around without intervention by a powerful leader. Many other factors may contribute to such turnarounds, but leadership is catalyst"

Studies conducted on effective schools in South Africa (Christie et al., 2007; NPC, 2011) show that even in poor areas a strong principal has a very large effect on school performance. Empirical research conducted in Soweto township (Mbokasi, 2015) has shown that successful schools are characterised by leadership that focuses on strategy, regulation, pedagogy, compensation and security of learners and staff. Similarly, Maringe, Masinire and Nkambule (2015) have highlighted that effective school principals in disadvantaged communities, adopt a service-oriented approach, work beyond the school timetable,

encourage staff stability and commitment, and make efforts to engage parents and stakeholders around a common school development project. Evidence has shown that effective principals can raise student achievement by 2-7 months in a year (Branch, Hanushek & Rivkin, 2012; Chiang, Lipscomb & Gill, 2012; Coellli & Green, 2012; Grissom, Kalogrides & Loeb, 2012). As turnover and replacement of school managers is extremely slow in South Africa, a principal, whether good or bad, sets the trajectory of a school for at least a decade (Wills, 2015).

Thus, to improve the quality of schooling systems, interventions are often made around strengthening school leadership. These usually consist of programmes for the training and coaching of principals and school management teams (SMTs). Such programmes can also be geared to putting new systems in place to improve school functionality, and promoting accountability, efficiency and good governance of schools at local level. These can include also the introduction of specific incentives and prizes for achievement and performance. Since 2007, the Department of Education has made sure that principals go through an accredited ACE School Leadership course and are also trained in various aspects of education management (Bush et al., 2009). Private sector, such as BHP Billiton and Old Mutual, have also undertaken similar trainings.

The IQMS, discussed earlier, does not rate school managers. However, the Education Amendment Act (2007) calls on increased accountability for principals, requiring them to produce a plan for academic improvement and report regularly against set targets. Other recent development from the DBE has been the introduction of Performance Management Rules and Draft Standards for Good Principals (DBE, 2011; Wills, 2015). The National Development Plan (2012), had announced the launch of performance contracts for principals and their deputies, but following negotiations within the Education Labour Relations Council, such performance management system for school management was withdrawn and left unconcluded (Wills, 2015).

Critical to the functioning of principals and school management teams, is also the support received by their School Governing Bodies (SGB) and the districts authorities (Taylor et al., 2013), an aspect which will be discussed further in the next section.

3.8.7 Systemic reforms and sub-national policies

As has been discussed until this point, the education system in South Africa is a complex bureaucracy that involves administrative layers of teachers, principals, school governing bodies, education officials at national and sub-national level. Therefore, to improve learning outcomes a number of different types of governance and accountability interventions need to be implemented at different levels of the system.

These are normally take the form of programmes and structural policies implemented by the National and Provincial Departments of Education.

Some of these systemic reforms will be analysed in more detail in chapter 5 meta-analysis, however it is worth presenting one of the notable interventions of the Gauteng Department of Education (GDE), aimed at improving learning outcomes in dysfunctional schools. In 1999 the Gauteng Education MEC announced the Education Action Zones (EAZ), a special programme which would provide provide 'intensified measure' (Fleisch, 2002) of teacher training and coaching, support to SGBs, weekly meetings with School Management Teams, monitoring of curriculum coverage and additional tutoring of learners in low performing schools of the province (Jacobs, 1999). The EAZ was managed directly by the office of the GDE Head of Department that provided an iron-fist approach to instate bureaucratic procedures, 'bullying' treats of disciplinary actions, incentives and pressures that would eventually lead to improvement of learning results, measured by matric pass rates (Dieltiens & Mandipaza, 2014; Fleisch, 2002)

In only a few years of implementation, the Education Action Zones caused a dramatic increase in pass rates of the Gauteng schools involved. Sapphire Secondary in the Vaal, for instance, went from 13,62% pass rate in 1999 to 77,7% in 2001 (Fleisch, 2002), however this was more as a result of 'gaming' by schools, who would deliberately keep back low performing learners in grade 11, thus producing better average results for the cohorts writing the NSC exams in grade 12 (Fleish, 2006). Taylor (2006) criticised the EAZ for being overly coercive and forcing accountability measures, but with limited support to the schools. Ntuta and Shurink (2010) explained that the EAZ was not sustainable, and once the MEC who had championed the initiative left office, SADTU put pressure on the Gauteng education authorities to stop the programme, and the schools went back to their previously low performing levels (Fleisch, 2006).

The majority of schools in South Africa still suffer from poor management capacity, thus procurement and recruitment processes are handled by the Provincial Departments of Education. The provincial institutions are however also weak, therefore this leads to bureaucratic delays and system-wide inefficiencies that often result in schools not receiving teaching equipment and material on time, as it occurred for instance with the Limpopo scandal in 2013 (Chisholm, 2013; Veriava, 2013), discussed earlier.

International studies (e.g., World Bank, 2011) have confirmed, in fact, that the more functions are centralised the more inefficiencies, bottlenecks and delays occur in the distribution of resources. Decentralisation therefore is very important for a well-functioning education system, as it empowers stakeholders at the local level to take more ownership and control of teaching and learning processes. We have seen earlier, that when schools are strong enough, the SASA (1996) encourages them to register as

'Section 21' institutions, which provides them more autonomy in mobilising resources, recruiting additional educators, procuring learning material and the management of school enhancement initiatives - thus leading to better learning results than other public schools (Fiske & Ladd, 2004; Woolman & Fleisch, 2006). Many programmes (discussed later in this research) are therefore geared towards the strengthening of school management and school governing boards (see section 3.8.9 and chapter 5).

An other important level in school governance are the district offices which provide the critical link between the provincial education departments and the schools - their management and their governing bodies. They are often further sub-divided into circuits or areas, for better management of the schools within the urban and rural districts. District and area offices consist of managers, inspectors, content advisors, which ensure LTSMs are delivered to the schools, teachers and principals are supported, and that curriculum delivery is on track (de Clercq, 2001; Mphahlele, 2002; Narsee, 2002). They are the closest governmental officials to the schools, and conduct regular monitoring and evaluation, ensures functionality of the education system in their area (Taylor et al., 2003). Prew (2002) explain that district officials are the pivotal and crucial nexus to large scale implementation and the sustainability of education reform programmes. Other research by JET (i.e., Roberts, 2001; Taylor & Prinsloo, 2005) and the World Bank (2011) have also highlighted the importance of district and other local education authorities. The pivotal role of district and circuit managers will be illustrated later also through the Anglo American case study in chapter 4.

Many efforts have therefore gone to also strengthen the capacity of sub-national education offices, including the establishment of the Planning and Delivery Oversight Unit (PDOU) in the DBE, to support district development. To improve learning results, particularly during the FET stage, subject advisors have also been introduced at district and circuit level to support teachers, assist with curriculum delivery and provide regular assessment tools. Fleisch (in Harris & Chrispeels, 2006) has explained that many District Development Programmes have been implemented in South Africa over the years including USAID's District Development Support Programme, JET's Mahlahle District Development Project, and the Kgatelopele District Improvement Project of the Open Society Foundation. Some of these will be discussed again in section 3.8.9.

3.8.8 Learner-targeted initiatives

At times when the public system fails to deliver, such as in the South African case with the underachievement of sufficient qualified mathematics and science school graduates, the private sector and non-governmental organisations step in to supply additional development interventions to disadvantaged learners within the schooling system.

In South Africa only pockets of research have been conducted in this field by CDE (2010b) and JET (Besharati, 2015). As was discussed earlier in section 3.7.3, in South Africa the private and non-governmental sector runs a number of parallel education programmes targeted directly to disadvantaged learners to support their academic empowerment. These programmes can be divided in two main categories: a) scholarships and learner placement programmes; b) supplementary tutoring and enrichment programmes.

a) Scholarships and learner placement programmes

Often in poor and disadvantaged communities talented and promising learners are found, but the social-economic conditions and the school environment are not conducive to their progress and development of their full potential. Foundations, NGOs, churches and private companies therefore run programmes to identify such talented learners and finance their placement in special course or schools which are known as centres of excellence. Sometimes the placement schools are institutions established by the funding agencies or sometimes they are simply the high performing public and private schools in the region. Some examples of such projects include LEAP Schools, Alan Gray Orbis Foundation, Make a Difference Foundation, Metropolitan, Student Sponsorship Programme, Sekolo sa Berogo, and Maths & English Programme.

Such programmes involve providing substantial scholarships for individual learners with talent from disadvantaged backgrounds and providing them financial support for school fees, learning materials, uniforms, transport, medical costs, meals and some times even boarding. Many of these programmes are residential therefore the scheme involves the relocation of the learner out of their home community into a new unfamiliar and challenging environment. For this reason, many of these programmes offer also emotional, psyco-social support and mentoring components. Cost of such schemes are fairly high and vary between 30,000 and 60,000 rand a year per learner (Taylor, 2011). Sustainability of such scholarship schemes throughout the schooling process can therefore become challenging.

Learner-targeted initiatives tend to be fairly expensive, and are often criticised for benefitting only a small group of individuals and not being sustainable (Besharati, 2014, 2015). CDE (2010b) has conducted a cost-benefit analysis of 10 prominent learner enrichment schemes offered by the private sector and has concluded that the most cost-effective programmes is the placement of disadvantaged talented children in the public 'Dinaledi' schools. However, the Dinaledi schools do not always have enough space for children from other areas as they need to prioritise learners from their geographic catchment area, according to the NEPA (1996) regulations. Impact evaluations of the Dinaledi programme will be further discussed in chapter 4 and 5 of this research.

b) Supplementary tutoring and enrichment programmes

These are typically implemented by private tutors or volunteers on a one-to-one basis or for a small to medium size group of learners to further consolidate the academic content imparted in the official school curriculum. These learner-based interventions are referred in the international literature as 'shadow education' (Bray & Lykins, 2012, Lee et al., 2009) as they mimic and run parallel to the mainstream education system. Shadow education has been growing throughout the developing world, but it has gained more prominence in Asia (Bray, 2011; Buchmann, 2002; Burch, 2009; Sobhy, 2012; Sujatha & Rani, 2011), where parental spending on shadow education almost equals government spending on public education (Bray & Lynkins, 2012). In South Korea 90% of elementary school pupils receive private tutoring, and similarly high figures are reported also for Hong Kong (85%) and India (60%) (Bray & Lynkins, 2012).

As a result of the general breakdown of the public schooling system, private tutoring and supplementary classes has been on the rise also in South Africa. These supplementary courses are offered not only by the private sector, but increasingly organised even by public sector institutions, such as district and circuit education authorities, municipalities and schools themselves; as will be seen also in the next chapter with the case study in Limpopo and North West province. These extra enrichment classes are often delivered by experienced teachers, in specialised subjects such as mathematics and science and are typically done after school hours, on week-ends or during the school holidays. Classes are usually held at the same home school or at facilities not too far from the communities. Learners are sometime offered transport support through bus and taxi services. Some examples of such projects include the Star Schools, Maths Centres, Science Incubator Programme, St. John's Academy, Roedean School and Waterkloof Hoerskool (CDE, 2010b). Companies which sponsor learner participation in such programmes include Investec, Telkom, Rand Merchant Bank, First Rand, BP, Anglo American to name a few. Costs of running such supplementary learner enrichment programmes vary from 5,000 to 7,000 rand a year per learner (CDE, 2010b). Sometimes such mathematics enrichment programmes are linked to provincial and national academic competitions such as the South African Maths Olympics.

Though these forms of supplementary tutoring contribute to human capital development and the provision of additional time and resources for disadvantaged learners with limited opportunities provided by the public school system, such initiatives could has also deleterious effects. Bray (2013) has explained how shadow education can undermine efforts towards equitable and inclusive access to education, and can question the relevance and cost-efficiency of the mainstream education system. Supplementary tutoring poses additional financial burden on parents and communities, increases stress and demands on

young people (who should be spending their time also in sports, social and other developmental activities), and can sometimes create perverse incentives and mild forms of corruption, as educators less time fulfilling their classroom teaching duties because they expect learners to attend their tutoring lessons after school (Bray & Lykins, 2012). As these supplementary classes occur mostly on an informal basis, tutors often do not require specific qualifications, and the shadow education industry continues to grow globally and in South Africa, without much regulation or quality control mechanisms (Bray, 2013).

3.8.9 Integrated school development models

As discussed in this chapter achieving improvement in the quality of schooling in South Africa requires a combination of many areas of interventions into a holistic model of school improvement (Mouton et al., 2014). In his historical account on school development in South Africa, Brahm Fleisch (2006) explains the evolution of these integrated education programmes:

"Drawing on the literature of whole-school development, these programmes focused on bottom-up development either through building collaborative organizational cultures at school level, or through structured processes associated with school development planning… as a results, a third wave of education improvement initiatives (Harvey & Peacock, 2001; Harvey, 2002) have emerged that focus at multi-levels, i.e. classroom, teacher, school management, and district… the assumption behind the third wave of education improvement initiatives is that improvement, in order to be at-scale and sustainable, must focus on consensus building, management improvement, curriculum and teaching improvement at all levels of the system."

(Harris & Chrispeels, 2006, p.219-22)

Over the past two decades, many institutions in South Africa - national and provincial governments, private sector and foreign donors - have come up with programmes that integrate several of the interventions (discussed in previous sections) into a comprehensive strategy to enhance school functionality and ultimately improve learning outcomes. These combined school development initiatives operate at different levels from classroom, to school, to districts, involving a range of different educational stakeholders in the transformative process. Such complex school development models are often conceptualised after years of research, based on special theoretical frameworks, tested in a few schools before they are brought to scale with large injections of funding. These large school development models are often implemented country-wide in hundreds of schools through partnerships between

government, private sector and specialised non-governmental agencies. Some examples of such school development models are for instance the ones developed and implemented by JET, Zenex Foundation, National Business Initiative, USAID, Limpopo and Gauteng Provincial Departments of Education, to name just a few.

Sayed et al. (2013) have provided a comprehensive review of several of these school development programmes, which have received over the years much debate within South Africa's academic community (Bloch, 2009; Fleisch, 2006; Mouton et al., 2014; Schollar, 2015). Many of these initiatives have been closely examined and have benefitted from rigorous impact assessments conducted on them (Blum et al., 2010; Fleisch & Schoer, 2014; Kanjee & Prinsloo, 2005; Schollar, 2005). Some of these programmes and evaluations will be discussed in more depth in the meta-analysis in chapter 5. However, for the time being, Table 6 provides a useful overview to the reader of some of the major whole school development programmes implemented in South Africa in the past 20 years, together with some of the evaluative and research studies conducted around these interventions. These programmes will be taken up again in subsequent chapters of this research.

Table 6. List of Major School Development Programmes Implemented in South Africa

Name of programme	Funders	Implementing Agency	Schools	Locations	Time-span	Total Costs	Studies & evaluations
Dinaledi Schools	DBE, Ireland, Various corporate	DBE	513 secondary	Country-wide	2001-2011	2 billion	World Bank (2010)
GPLMS	GDE	various NGOs	811 primary	Gauteng	2010-2014	1 billion	Fleisch & Schoer, 2014
School Development Project	Zenex	Various NGOs	200 mix	KZN, Gauteng, WC	2006-2010	200 million	Mouton, 2012
Education Programing	Anglo American group Epoch & Optima Trust	Various NGOs & Gov.	mix	Limpopo, NW, Mpumalanga	2005-2011	2,5 billion	Schollar, 1999
Learning for Living	Business Trust	READ	967 primary	Country-wide	2000-2004	260 million	Schollar (2005)
Quality of Learning	Business Trust	JET	524 secondary	Country-wide	2000-2004	140 million	Kanjee & Prinsloo (2005)
EQUIP	Various corporations	NBI	500 mix	Eastern Cape, Gauteng, KZN, WC	1995-2008		Schollar, 2002
Kanyisa School Transformation Model	UK-DFID	JET	1000 primary	Limpopo	2002-2009	130 million	Simkins & Pereira (2010)
Child Friendly Schools	UNICEF	Various NGOs	616 mix	Eastern Cape, Limpopo, KZN, NC	2007-2010		UNICEF (2011)
DDSP	USAID	RTI + Various NGOs	589 Mix	Eastern Cape, NW, KZN, NC, Limpopo	1998-2003	$150 million	HSRC (2003), Schollar, 2006
PrimED SBSP	European Union / Netherlans	DBE, Provincial DE		Limpopo - country-wide	2006-2013	1,5 billion	EU-SA SBS evaluation (2013)
Education Action Zones	GDE	GDE	70 secondary	Gauteng	1999-2001		Fleisch, 2006
PMRP	Various	ESA	125 primary	Limpopo	2007-2012		Schollar, 2012. Fleisch et al., 2010
Imbewu I & II	UK-DFID	JET + other NGOs	700 primary	Eastern Cape	1996-2007	7,5 million GBP	Roberts, 2000, Schollar, 2000

Source: Author's own compilation (2016)

3.9 A theory of change for South Africa's education sector

This chapter has provided a brief overview of the main socio-economic challenges, institutional frameworks and policy architecture put in place in South Africa's education sector post-1994. It has highlighted the dire problem of poor learning outcomes which still persist in South African schools, and the investments which have been made by the government, private sector, civil society and other partners, to address the quality of education in South Africa. The chapter has provided a review of some of the main types of interventions implemented by different institutions to address the short-falls in learner achievement in South Africa's school system.

Improving the quality of basic education in South Africa will increase the country's global competitiveness, and allow for a new class of educated and skilled young Africans who will drive economic growth, affect socio-political transformations, and reduce the poverty and inequality gaps, still prevalent in the nation. The interventions in South Africa's schooling sector have ranged from upgrading of infrastructure and facilities, provision of learning and study materials (LTSM), teacher development programmes, learner-focused enrichment, strengthening of school management, putting in place accountability systems, initiatives to engage parents and communities, strengthening different parts of the education system, structural reform and whole school development programmes. Numerous programmes and policies, capturing a variety of these elements, have been implemented by national and sub-national government entities, corporate investors, foreign donors and a range of NGOs and service providers.

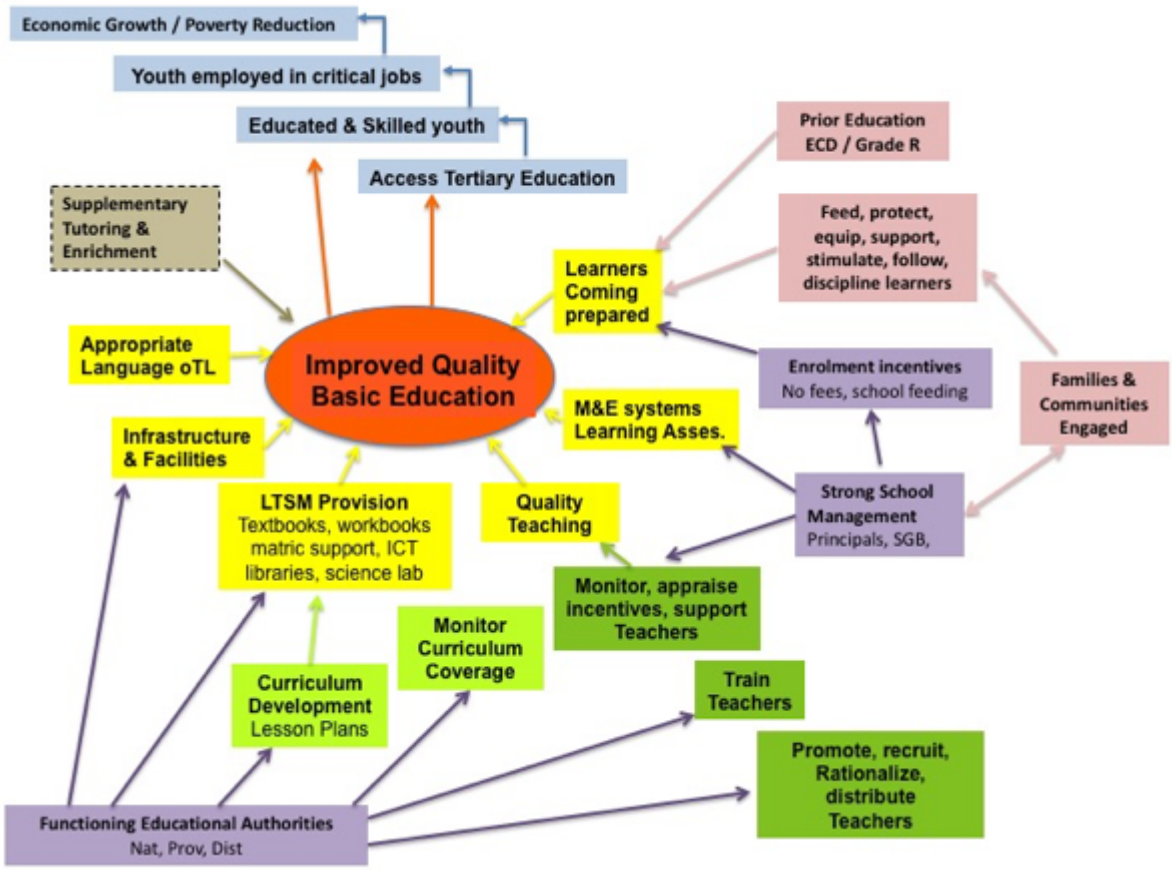
None of the above interventions however operate in a vacuum. The education system is a complex apparatus, and learner achievement usually comes about as a combination of different forces, from the family and community, to the school system, and from inputs and support received from different interventions and different external institutions. As discussed in chapter two of this research, before analysing further the effectiveness of any development intervention, the theory of change that underpins the sector of operation needs to be carefully unpacked. This process allows for the examination of the inner functioning of the systems and the dynamics of how the programmes and policies work within a particular developmental context.

This chapter has thus reviewed the existing literature on South Africa's education field and the programmes and policies carried out throughout the two decades of democratic dispensation to address the challenges of the sector. The contextual background provided in this chapter helps understand the policy environment, the socio-economic factors, institutional forces, and the theories behind the interventions, which will be analysed in more detail in the empirical work in chapter 4 and 5. The mapping out of all the factors that influence the outcome of interest (learner achievement) become

important also when conducting the qualitative field work, the econometric analysis and the quasi-experimental evaluations undertaken in the next two chapters of this research endeavour.

To assist in providing the reader a graphical overview of the theory of change operating in South Africa's education sector, and which guide many of the interventions discussed in the above chapter, the following logic model (Figure 23) has been drawn up to show how the various education policies and programmes interact with one another and with other social forces, in order to arrive at the expected outcome of improved quality of South Africa's schooling sector. Figure 23 shows that to achieve the development outcomes (in blu) of economic growth and poverty reduction in South Africa, youth need to be appropriately skilled and employed in critical jobs required by the economy. For young people to be able to enter tertiary education and training, there needs to be a strong primary and secondary education system in the country. The elements required for a functioning basic education system (yellow boxes) include learners coming prepared to well-resourced schools with appropriate infrastructure, facilitaties and LTSM, accompanied by high quality teaching in the right language, and with the appropriate forms of assessment, monitoring and accountability mechanisms. The later can be facilitated by a strong school management (purple boxes), backed up by families and communities, that ensure that learners receive adequate support at home and before coming to school (pink boxes). For quality teaching to occur, there needs to be enough teachers in the system, appropriately trained, supervised, incentivised (dark green boxes), and supported with a solid curriculum framework, whose proper implementation is carefully monitored by circuit, district and provincial authorities (purple boxes).

Figure 23. Theory of change for the operation of South Africa's education sector.



Source: Author's own compilation

If all of these elements in the diagram above would work well and in tandem, the envisaged development outcomes would be achieved, however the irony is that in South Africa, each and everyone of the above components has severe problems which has been discussed extensively in the literature (Bloch, 2009; Chisholm, 2004; Fiske & Ladd, 2004; Fleisch, 2008; Motala et al., 2014; Sayed et al., 2013; Spaull, 2013; Taylor et al., 2013). When the public system fails, as in the case of South Africa, the private sector often steps in with parallel education, supplementary tutoring and enrichment programmes (Bray & Lynkins, 2012), as illustrated in the gray box in Figure 23.

By no means this is a definitive theoretical framework for South Africa's education sector; it simply is a visual summary of the theory of change that the researcher has gathered from the South African academic and policy literature on education, which has been surveyed throughout this chapter. This framework is for the general system; however, every education intervention has its own theory of

change, some of which are elaborated in more detail in chapter 4 and 5, when the specific programmes and policies are being discussed.

What is more surprising is that with the numerous forms of interventions and large volumes of investments made by both public and private actors in South Africa over the past 20 years (see section 3.7), the learning outcomes in South African remain dismally poor, and below the standards of even other lower income countries in Africa (Spaull, 2013). Section 3.6 of this chapter has provided a more detailed account of some of the results from the international assessments that South Africa participates in, which have been summarised by The Economist (3 June, 2010):

"South Africa spends 6,1% of its GDP on education, a bigger chunk than most other countries, yet its results are among the worst. In the World Economic Forum's latest Global Competitiveness Index, it ranks bottom (out of 133 countries) in both maths and science education. In the 2006 Progress in International Reading and Literacy Study it also came bottom (out of 40 countries), as it did in 2003 Trends in International Mathematics and Science Study (out of 48 countries)"

The vast array of complex interventions to improve the South African education system, have unfortunately not translated into the desired results hoped for by policy-makers, development managers and investors. Notwithstanding the volumes of education research conducted post-1994 in South Africa, there is still limited information on the successes and failures of school development interventions, and the lessons therefrom that can be extracted for South African education policy (Sayed et al., 2013).

"Since the early 1990s we have seen a plethora of both governmental and non-governmental activity in schooling reform, which have taken the form of both programme interventions for improvement, and research into the nature of schooling and the effect of programme interventions. It is surely too much to say that we are no wiser at all, but only the foolhardy will claim that we have any firm answers"

(Taylor, Muller & Vinjevold, 2003, p. 128)

With all the studies conducted in the sector over the decades, scholars concur that the evidence is still too weak to draw any conclusions on what works and what doesn't work in South Africa's education system (Bloch, 2009; Taylor et al., 2003). When commenting on education evaluations in South Africa, Motala and Pampallis (2005, p. 61) explain that "the literature is small in scope, and in general lacks the methodological and conceptual rigour. Much of the valuable work is written in the form of policy briefs and observations are not based on in-depth, sustained research".

It is in within the knowledge gaps in this important development policy sector for South Africa, that this research tries to position itself and to answer important questions on the minds of education practitioners and evaluators. By utilising the evaluation framework, the approaches, techniques and the rigorous methods discussed in chapter 2, the research hopes to shed more empirical evidence on the programmes and policies aimed at improving learning outcomes in South Africa's schooling sector, and identify some of the most effective interventions that can be considered for replication and upscaling.

The conceptual and methodological framework discussed in this thesis, will be piloted on two empirical studies within South Africa's education sector. The first will be a rigorous impact evaluation of the education programme of Anglo American, the biggest corporate social investor in South Africa, implemented in public schools in Limpopo and North West province. The second will be a systematic review and comparative meta-analysis of some of the major interventions implemented in the past two decades to improve language, mathematics and science learning outcomes in the country. These will be dealt with respectively in chapter 4 and chapter 5. Through these case studies more learning can occur about the key features, factors and context that contribute to the success of development interventions aimed at improving the overall quality and delivery of South Africa's education system.

Chapter 4: Case Study:

Evaluating the impact of Amplats education programmes in Limpopo and North West province

4.1 Introduction

Chapter 2 presented a review of some of the major evaluation approaches, methods, and tools utilised to measure and compare effectiveness of development interventions. This served as the theoretical foundation for the development of the conceptual and methodological framework proposed in this research. To illustrate the proposed set of analytical approaches and tools, the education sector of South Africa was chosen as a first case study that presented many of the right conditions for the application of the evaluation framework (see section 1.5). Chapter 3 provided a contextual overview of the historical and institutional developments in South Africa's education sector (Fiske & Ladd, 2004), the policy environment (Motala et al., 2014; Woolman & Fleisch, 2008) and the diversity of public and private stakeholders (Besharati, 2015), who have invested substantially in a range of diverse interventions to improve the poor learning outcomes (Spaull, 2013), still persisting in public schools (Sayed et al., 2013; Taylor et al., 2013)

One of the first development initiatives which was empirically analysed, through the methodological framework proposed in this research, was the 100 million rand education programme of Anglo American Platinum, one of the biggest corporate social investors in South Africa (Trialogue, 2014). As part of its support for the research endeavour, Anglo Platinum Community Engagement and Development (CED) unit, agreed to provide funding and facilitate access to data for the lead researcher to undertake a thorough impact assessment of its education programme, and thus receive from the scientific exercise useful evidence to inform its future investments in community development activities around its mining areas. The following chapter is thus based on an evaluation conducted by the same author of this research (Besharati, 2014) for a private client, Anglo American Platinum, who agreed to make available the data and findings of the study also for academic purposes.

The study explores the impact of the Anglo Platinum education interventions implemented in Limpopo and North West province, aimed at improving learning outcomes in public schools, particularly in the critical subjects of mathematics and sciences. This pilot study illustrates with a real-life example the application of the evaluation framework, discussed in chapter 2. As part of the set of methods and approaches proposed in this research, the following study utilises a theory-based impact evaluation approach, which combines quasi-experimental techniques, econometric and meta-analytical tools with traditional qualitative and participatory methods of enquiry.

The study included extensive document review as well as interviews with stakeholders during field visits undertaken from September to December 2013 in the two rural provinces of Northern South Africa, to better understand the schooling context, the theory of change underlying the interventions, and the other socio-economic and institutional factors influencing the development outcome of interest (which also served as control variables during the evaluation). The study utilised a rich pool of mixed data to conduct descriptive and inferential statistics, while also engaging a large set of institutions in a participatory manner during the evaluation process and in the reflection on the findings.

The research reveals some surprising findings with regard to the effects that mines have on learning results of surrounding schools and affected communities. The case study also re-affirms much of the South African education literature discussed in chapter 3, and opens new questions with regard to the role of educators, the impact of interventions, socio-economic factors, and other aspects affecting the school system. Moreover, the chapter illustrates some of the problems and shortfalls of agency evaluations discussed earlier in section 2.8. As will be seen in section 4.7.2, findings of the case study provided insights and sparked debate on the effectiveness of interventions, not only of Anglo American, but of the broader private and public sector, concerned with addressing education and development challenges in South Africa.

The chapter commences by providing a contextual background about the mining industry in South Africa's development processes, the state of learning outcomes in schools in the specific provinces of Limpopo and North West, and an overview of the programmes implemented by Anglo Platinum to improve mathematics and science learning outcomes in targeted schools. It will subsequently elaborate on the data preparation and the methods and approaches used for the qualitative field work as well as the quantitative impact evaluation. The last section will explain the results emerging from the field interviews, the econometric and quasi-experimental evaluations, and conclude by discussing the implications of the findings of the case study for corporate social investors, for public education policy, for the field of development evaluation, and for future research in these domains.

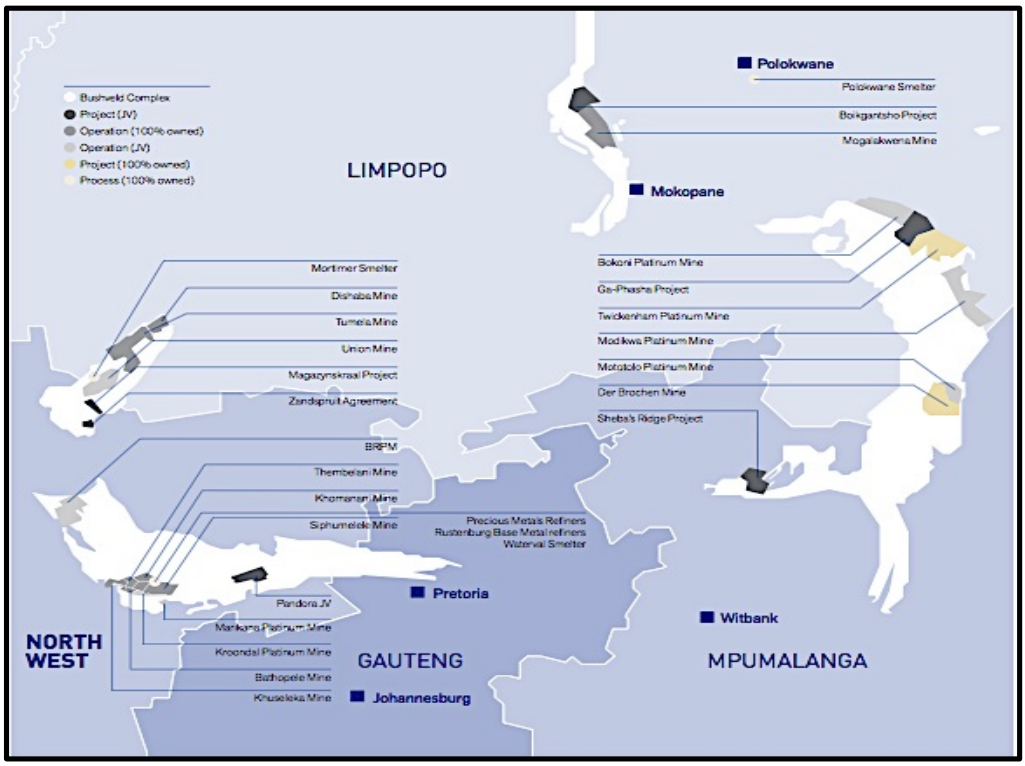
4.2 Contextual background of the case study

4.2.1 The mining sector in South Africa's development

Mining has played a central role in South Africa's history, economy and social structures. Migration to urban areas, the wide divide between capital and labour, mostly on racial lines, and the extreme social inequalities that fuelled the political system of apartheid are all intrinsically linked to the mining industry (Johnstone, 1976). Even the rise of the metropolis of Johannesburg and the township of Soweto, was driven on the back of the 'Witwatersrand Gold Rush' of the late 1800s (Callinicos, 1993). Less dominant than in the previous century, today the mining sector generates up to 18% of the nation's GDP and provides direct employment to over 500,000 people and indirectly another one million South Africans (Chamber of Mines, 2013).

South Africa contains large deposits of numerous minerals from chromium, iron ore, diamonds, palladium, gold to coal from which most of the country's energy is produced (Chamber of Mines, 2013). But since 1990, platinum has emerged as the largest component of South Africa's mining sector (Ashman, 2013) with around 130 tonnes of output a year, which grew by 67% from 1994 to 2009 (Capps, 2013). South Africa holds 80-90% of the world's platinum reserves (Stilwell & Minnitt, 2006). The Merensky Platinum Reef stretches from Southern Zimbabwe through Polokwane all the way to the Rustenburg region, with the largest concentration around the Bushveld Complex in the Bojanala district, which since the early 1900s houses some of the major platinum corporations such as Anglo American, Impala, Lonmin, and Bafokeng Rasimone Mines. Another Platinum-rich area is the Twickenham Reef in Sekhukhune district, situated roughly 100 km south-east of Polokwane. The map below (Figure 24) illustrates the location of some of the major platinum extraction sites in the Limpopo and North-West provinces:

Figure 24. Locations of platinum deposits in Northern South Africa.



Source: Anglo Platinum[2] (2013)

South African platinum production is mainly geared towards international exports which have been growing since early 2000 (Capps, 2013), driven by demand from the European automotive, medical, and chemical industry (Ashman, 2013) and by the growing Chinese jewellery manufacturers (Benkenstein, 2013). Nonetheless, due to the increasing cost of labour and electricity in South Africa (Chambers of Mines, 2013), platinum prices have remained stagnant, with expansion projects put on hold and potential threat of mine closures. The labour situation has been badly deteriorating during the course of 2012 and 2013, with massive retrenchments, on-going strikes, workers unrest, union fragmentation, violence, and even murders leading to the notorious 16 August 2012 'Marikana massacre' which made international headlines, with 34 miners losing their lives in clashes with the police.

Like the rest of the mining industry, the platinum houses of South Africa are regulated by the Mineral and Petroleum Resources Development Act (MPRDA, 2002) which is aimed at minimising economic turbulence and depletion of natural resources. The Act also makes provisions for human resource issues such as employment equity, health, work safety and living conditions for mine workers. The Department of Mineral Resources (DMR) provides licences and rights for mineral exploration and production to private companies, based on certain social-economic-environmental conditions that need to

[2] http://www.angloplatinum.com/sus/overview/operations.asp

be fulfilled in and around mining operations (Capps, 2013). Intrinsic part of these requirements is the development of SLPs detailing how the companies deal with human resources (employment equity, retrenchment, housing, skills development) and support to the social, economic, and infrastructure development of the communities around the mines. These SLPs are generally developed in a participatory manner, through ethnographic techniques and through engaging the numerous stakeholders of the area such as community groups, local government, traditional authorities, labour unions, NGOs, and other civic groups, through consultative forums. Social and Labour Plans are generally supposed to be aligned to the Integrated Development Plans (IDPs) of the specific locality (MPRDA, 2002).

Another impetus to aligning the mining sector to the country's transformation process was the introduction by the Department of Trade and Industry (the DTI) of the Black Economic Empowerment (BEE Act, 2003) aimed at transferring 25% of the country's productive assets to black ownership, by transforming shareholders, management, staff, skills development, procurement of sub-contractors in favour of historically disadvantaged South Africans (HDSA). BEE compliance was a primary determinant to whether companies were granted conversation from the old to the new order rights under the MPRDA (2002). Within the BEE framework, the Mining Charter (developed in 2004 and updated in 2010) would also provide higher score cards to companies that would ensure sustainable environmental management, facilitate local beneficiation of commodities and resources, and support local community development activities.

The last decade saw the intensification of 'sustainability reports' by South Africa's private sector to highlight companies' contributions to the country's and community's welfare (Besharati, 2015). This has resulted in major mining houses such as Anglo Platinum, Kumba, Impala, De Beers, BHP Billiton spending over R100 million every year in CSI (Trialogue, 2010, 2013). Community investments by mining corporations in South Africa are implemented for a mixture of reasons and motivations, including securing social licences, maintaining good relations with local stakeholders, improving BEE score cards, marketing and enhancing public image, or simple and genuine philanthropic efforts aimed at making a positive change in society (Besharati, 2015). In recent years CSI approaches have also been more pragmatic, driven by 'enlightened self-interest', stemming from a realisation that having a healthy, educated, uplifted, and prosperous local community is ultimately also 'good for business' (Hamann, 2004). By the same token almost 92% of corporations in South Africa invest in education, because a knowledgeable and skilled population will enhance the workforce, the clientele, and the quality of the service providers of the private sector, and thus produce long-term returns for business. Overall the local corporate sector financing to South Africa's education sector surpasses by and large ODA from all the traditional bilateral and multilateral donors (Besharati, 2013b). Just one company such as Anglo

American spends more every year in social and education projects than the likes of the World Bank, the United Nations, JICA, CIDA, UK-DFID, and other foreign aid agencies operating in South Africa.

4.2.2 The state of the schooling sector in Limpopo and North-West

As previously discussed in section 3.7.3 (see also Besharati, 2013b, 2015), the foundation of the South African economy is the mining and the service sector (financial, logistical, telecommunications, retail). Ironically the country is not able to meet the high demand for engineers, financial analysts, and technical specialists required by the key sectors to run the economy and allow for further growth. There are not enough young South Africans entering and completing University degrees in such technical fields, greatly due to very poor performance of the South African secondary school education system, which is not able to produce enough learners with the minimum competencies in mathematics and sciences (Bloch, 2009; Fleisch, 2008; Spaull, 2013). Roughly 4% of South African learners who start grade 1 finish grade 12 with a pass rate above 40% in mathematics (Hodgson, January 2014, personal conversation). An older study (Schollar, 2008) explained that of the total number of learners enrolling in grade 1 in the new post-1994 South African education dispensation only 1,5% achieved a pass in HG mathematics during their 2006 grade 12 SC examinations.

This poses a serious challenge for the mining industry that needs to invest in education, not only for philanthropic reasons, but for practical reasons in order to increase and improve their labour pool of future technicians, artisans and engineers. Due to the national socio-economic transformation processes discussed in the previous section, mining companies have been giving special attention to sourcing future employees from around their areas of operation, focusing on the black population and the indigenous communities affected by the mining. Yet companies such as Anglo Platinum have reported that they struggle to award much of their available university scholarships and bursaries to youth from the communities around their operations (Mabe, 2013).

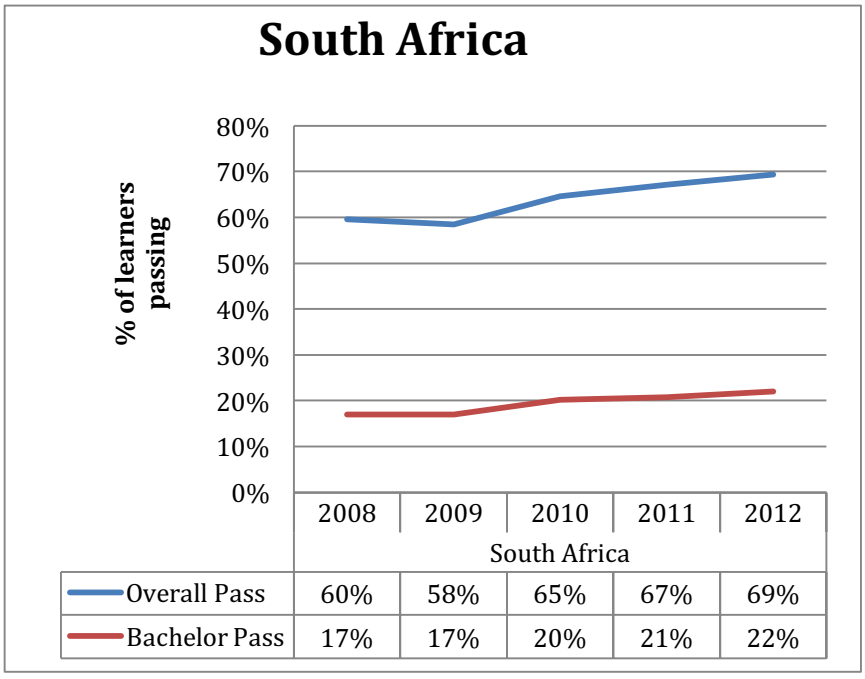
It would thus be appropriate to start looking at the academic performance of schools in the areas where the platinum mining operations occur, namely Limpopo and North-West province, and at sub-sets such as the Bojanala district. As discussed in previous section 3.5, the results from the NSC examination, undertaken at the end of grade 12 in all South African secondary schools, are potentially the best indicator to measure learner achievement and progress of schools and regions (Kanjee, 2007; Reddy, 2006). NSC exams are independently assessed by the central agency of Umalusi, using a universal system. Data is checked for quality and made available annually at micro-level.

As pointed out also in section 3.5, the downfall of using the NSC examination results as the main indicator for progress in learning is that external interventions, schooling, and social factors have less of

an impact on older and formed pupils rather than younger children. Bloom et al. (2008) have for instance shown that larger percentage of learning gains occur at lower grades, rather than at the end of the formal schooling process.

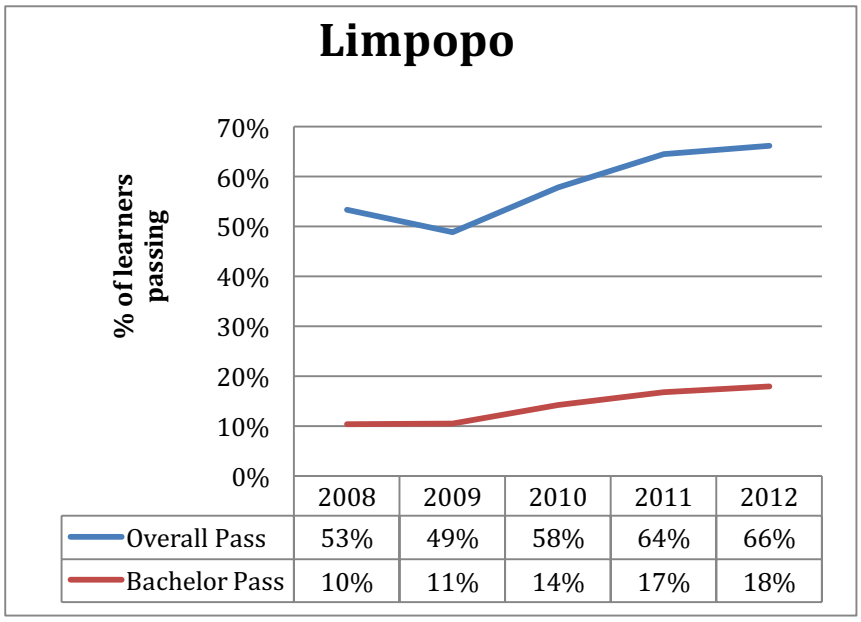
Nonetheless, the NSC can still provide in this case a good indication of South African learner 'preparedness' for university (Rankin, Schoer, Sebastiao, & van Walbeek, 2012) and training in commerce, technical and engineering subjects, most needed by the South African economy. Aside from overall NSC pass rates, of particular interest are the *bachelor passes* that allow for access into university. As the NSC examination system changed between 2007 and 2008, the graphs in Figure 25 have tried to capture the progress in the overall and bachelor pass rates in the Platinum regions of concern from 2008 till 2012, in relationship to overall country average.

Figure 25. Overall and Bachelor pass rates for NSC examination in South Africa.



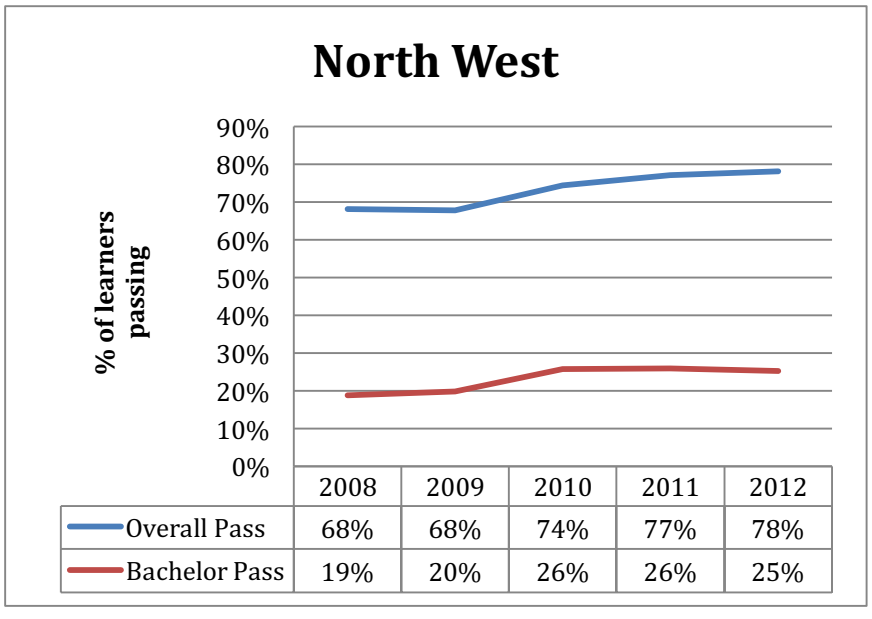
Source: Authors own compilation (2013)

Figure 26. Overall and Bachelor pass rates for NSC examination in Limpopo



Source: Authors own compilation (2013)

Figure 27. Overall and Bachelor pass rates for NSC examination in North West



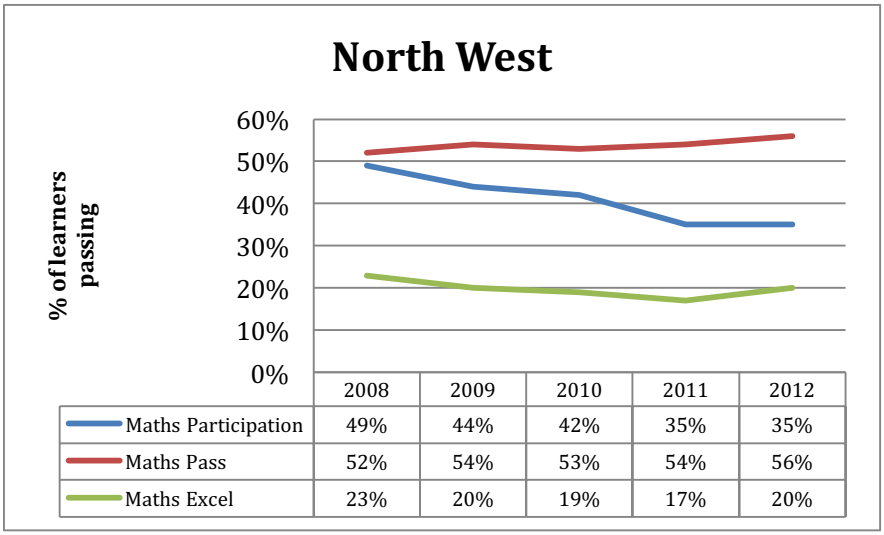
Source: Authors own compilation (2013)

In Figure 26 above, it can be observed that schools in North-West province tend to perform above the national average, while the schools in Limpopo (Figure 27) generally perform slightly below the national standards, most likely due to the different socio-economic rural conditions which differentiate these two South African provinces. Also due to the presence of mining operations in Bojanala district, North West receives more attention and initiatives by various public and private institutions. Limpopo province manages three times more schools than North-West province (DBE, schools master list, 2012), thus having a thinner spread of educational resources to distribute among its constituents.

The next graphs focus on the special 'gateway subjects' of mathematics and physical sciences, due to their importance for the economy and the mining industry. Again close attention will be given to the subject passes above 50% (excel passes), required for entry into most South African university; although Hunt et al. (2011) have shown how with the change in the new NSC system in 2008, grades above 60% in mathematics predict more accurately satisfactory performance of university students in the commerce and scientific subjects.

As explained in section 3.5, to pass the grade 12 exams, a minimum score of 30% (not 50% like in other countries) is sufficient in specific subject areas, thus highlighting the general poor quality of the education system of South Africa. One must also note that NSC pass rates are also masked by the fact that many South African learners drop out of school before they ever get to write the exam (Fleisch, 2006; Kanjee, 2007; Spaull, 2013). To give an indication of the severity of such phenomena, the number of learners in grade 12 are on average half of the number of learners who enrol in grade 10 in most South African schools (DBE, 2010). In the graphs below, mathematics participation and science participation relates to the percentage of learners writing the NSC who chose to take those specific subjects as part of their NSC examination.

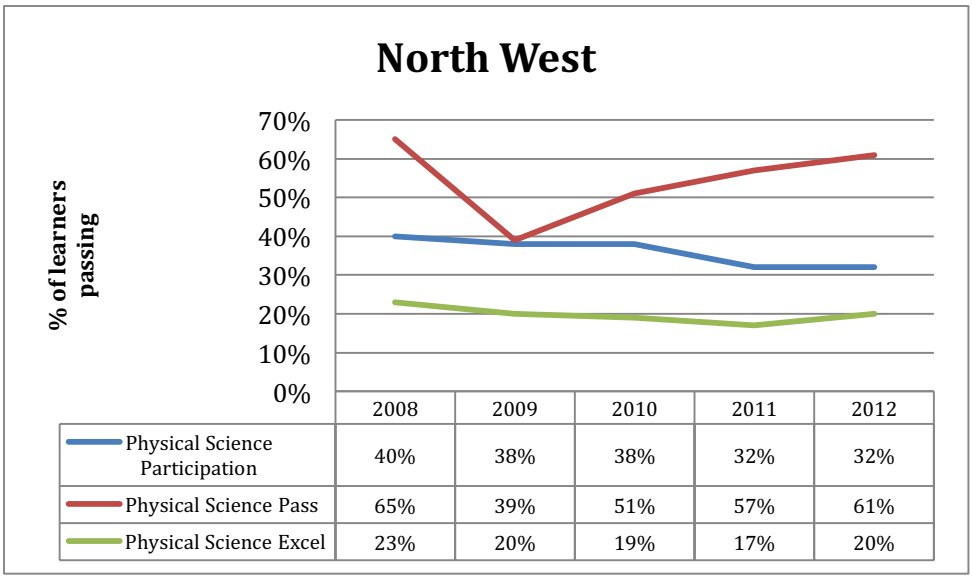
Figure 28 Maths pass rates for NSC examination in North West.



	2008	2009	2010	2011	2012
Maths Participation	49%	44%	42%	35%	35%
Maths Pass	52%	54%	53%	54%	56%
Maths Excel	23%	20%	19%	17%	20%

Source: Authors own compilation (2013)

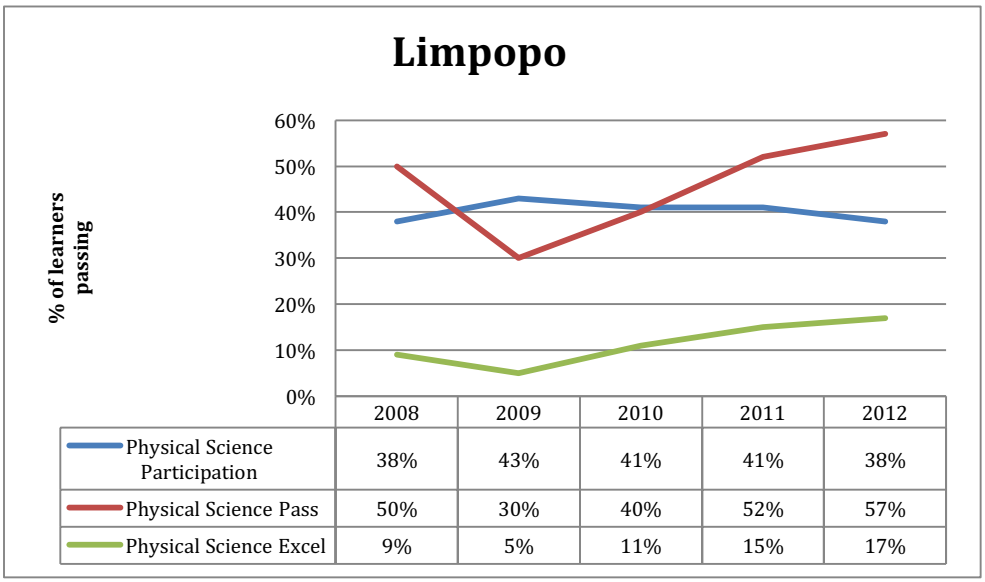
Figure 29 Physics pass rates for NSC examination in North West.



	2008	2009	2010	2011	2012
Physical Science Participation	40%	38%	38%	32%	32%
Physical Science Pass	65%	39%	51%	57%	61%
Physical Science Excel	23%	20%	19%	17%	20%

Source: Authors own compilation (2013)

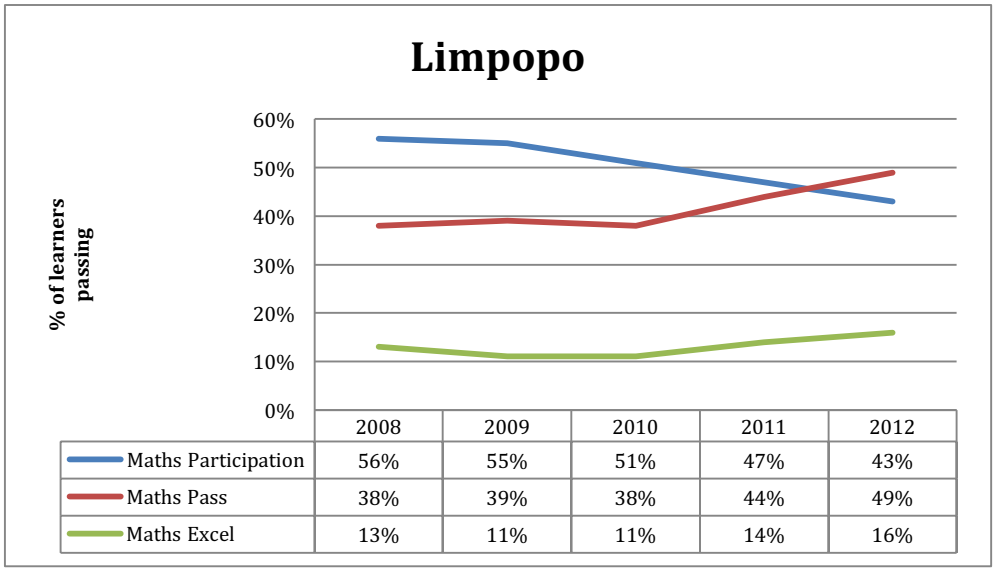
Figure 30 Maths pass rates for NSC examination in Limpopo.



	2008	2009	2010	2011	2012
Physical Science Participation	38%	43%	41%	41%	38%
Physical Science Pass	50%	30%	40%	52%	57%
Physical Science Excel	9%	5%	11%	15%	17%

Source: Authors own compilation (2013)

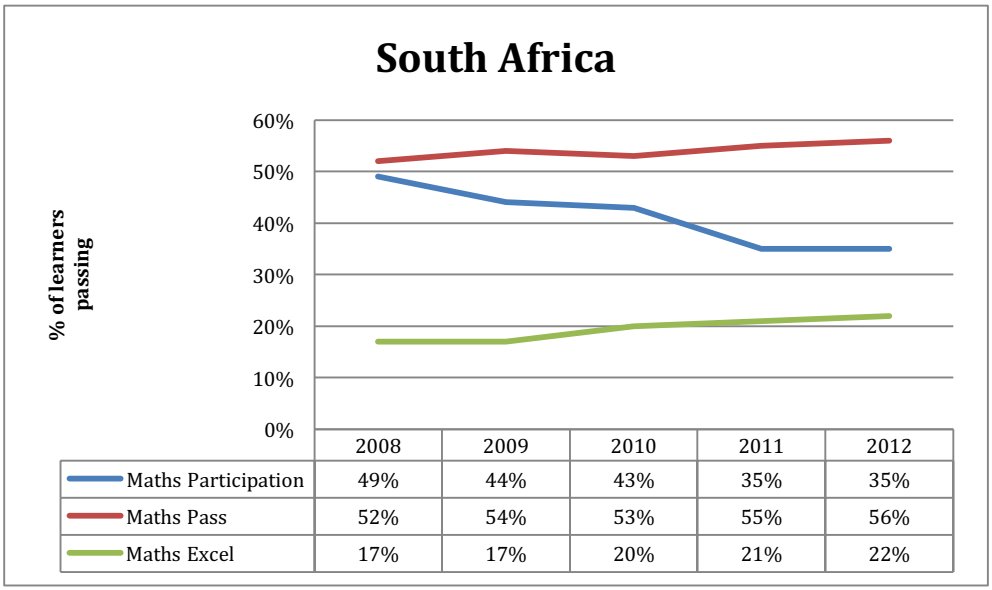
Figure 31 Physics pass rates for NSC examination in Limpopo.



	2008	2009	2010	2011	2012
Maths Participation	56%	55%	51%	47%	43%
Maths Pass	38%	39%	38%	44%	49%
Maths Excel	13%	11%	11%	14%	16%

Source: Authors own compilation (2013)

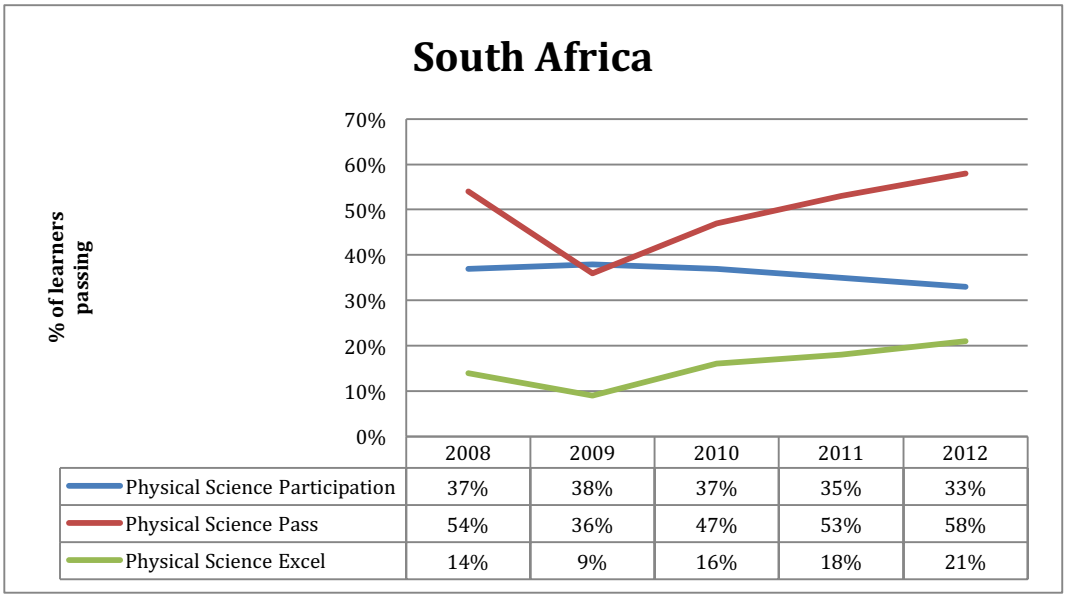
Figure 32 Maths pass rates for NSC examination in South Africa.



	2008	2009	2010	2011	2012
Maths Participation	49%	44%	43%	35%	35%
Maths Pass	52%	54%	53%	55%	56%
Maths Excel	17%	17%	20%	21%	22%

Source: Authors own compilation (2013)

Figure 33 Physics pass rates for NSC examination in South Africa.



	2008	2009	2010	2011	2012
Physical Science Participation	37%	38%	37%	35%	33%
Physical Science Pass	54%	36%	47%	53%	58%
Physical Science Excel	14%	9%	16%	18%	21%

Source: Authors own compilation (2013)

As seen from the graphs (Figures 28-33), over the years there has been a gradual and expected increase in the general pass rates as well as in the bachelor passes. In early 2011 Minister of Basic Education, Angie Motshekga, announced that despite the 'lost time' in 2010 due to the Football World Cup and extensive teachers and learners strikes that occurred that year, pass rates across the country received a drastic and surprising spike (DBE, 2011). The big dip in 2009 science passes was confirmed by Umalusi and explained by the fact that the science exam paper for that year was particularly difficult thus resulting in a drop of science passes across the country (Sibanda, 2013). An interesting correlation can be observed between a gradual decrease in the number of learners taking mathematics and science subjects at the NSC exams and an increase in the pass rates in those same subjects. This is possibly due to tighter learner selectivity, smaller teacher-to-learner ratios, and better spread of learning resources.

All of these are important trends that need to be borne in mind before undertaking further impact assessments on any specific intervention implemented to improve mathematics and science learning outcomes in the two provinces. This is to ensure that general socio-economic, environmental and schooling trends do not confine or mislead the results of the impact evaluation (Johnson & Lamdany, 2005; Mouton et al., 2013; White, 2013).

4.3 The Amplats education programme, its theory and evaluation

The following section provides a descriptive overview of the specific development intervention under analysis in this case study, namely the education investments of Anglo Platinum corporation in the schools around its mining sites. The following background is based on a review of reports, evaluations, and organisational documents, as well as researcher observation and interviews with company representatives, and other stakeholders (government institutions, community representatives, service providers, and civil society organisations) engaged in Anglo Platinum's education programme. It unpacks the different elements and features of the CSI intervention, and the theory of change that underpins the various activities. The following descriptive analysis reflects the preliminary steps in the evaluative exercise of the methodological approach discussed earlier in chapter 2. The subsequent more complex steps in the evaluation process, will be explained in more detail in the next section 4.4, which will elaborate on how the impact evaluation framework will be applied to answer the specific questions raised in this case study, so to produce empirical evidence for policy-makers, investors, and sector stakeholders.

There are several big mining houses in South Africa that undertake extensive CSI programmes in education (Trialogue, 2015), among which are Kumba, Impala, BHP Billiton, De Beers, and Rio Tinto.

This study however focuses on the education interventions of Anglo Platinum (Amplats) implemented in North West and Limpopo between 2009 and 2012. By far Amplats is the largest platinum producer in South Africa (Africa Mining IQ, 2016), producing 40% of the world's platinum output and employing over 50,000 people every year (Anglo Platinum Sustainability Report, 2012). Previously known as Johannesburg Consolidated Investment (JCI), the company has been undertaking education interventions since 1990s as revealed by a 1999 report of Eric Schollar and Associates (ESA) evaluating Amplats school development programme, whose features were not drastically different to what it is today. Together with Kumba, Anglo Platinum manages the largest budget of CSI of the Anglo American group (Anglo American, 2012). Within its Corporate Affairs division, the company established a sub-unit called Community Engagement and Development (CED) that deals with the mines' multiple stakeholders and develops, finances, implements, and monitors social-economic development programmes by Amplats in favour of local communities where it mines.

The Amplats education programme was based on its 2010-2014 SLP which, according to the company, was negotiated with the communities, local government, and other stakeholders affected by its mining operations in Magalakwena, Amandelbult, Union, Rustenburg, Twickenham, and Der Brochen. These included also local tribal groups present in North West and Limpopo such as the Bafokeng, Bengwenyama, and Langa Mapela. Some of the labour sending areas in the Eastern Cape, such as OR Tambo District Municipality and Taung community, were also part of Anglo Platinum's SLPs. Anglo American Platinum's SLP (2010) outlined community development activities in areas of health and welfare, poverty reduction, agriculture development, job creation, and establishment of basic community infrastructure. Within the SLP (2010), education and skills development, initiatives played also a significant role.

Shortly after finalising its SLP, Anglo Platinum pressed ahead to sign a Memorandum of Understanding with both North West as well as Limpopo Department of Education, through the Member of the Executive Committee (MEC) or the Head of Department (HoD), in order to implement its whole school development programme. The researcher also observed that as the programme unfolded, Amplats local CED officials would regularly liaise with district and circuit managers closer to the areas affected by the mines, and would collaborate closely in the implementation of the Anglo Platinum's support activities to local schools.

According to company and beneficiary reports, a major part of Amplat's infrastructure support has been spent in building or rehabilitating schools, extending classrooms, constructing administration blocks, toilets, water and sewerage facilities for education centres close to the mining areas. This has also included refurbishing libraries, providing science labs, donating computers, photocopier and other equipment to schools and even to provincial department circuit offices. According to the Anglo Platinum

CED annual reports (2009-2014) over the 5-year period, the company has invested over R25 million worth of infrastructure initiatives in schools around the mines. Anglo Platinum has further budgeted nearly R70 million for education, skills development and youth-targeted initiatives during the implementation of its SLP (2010). The education programme of Amplats has covered many areas including adult education and literacy (ABET), early childhood development (ECD), Primary and General Education Training (GET), special needs schools, bursaries for university students, apprenticeships, technical and professional development for young professionals.

This current case study, focuses on Anglo Platinum's Further Education and Training (FET) programme done in secondary schools, which has as its primary objective to improve learner results, particularly in mathematics and science, for which graduate throughput in those areas is particularly important for the future engineers and technicians required by the South African mining industry (see previous discussions in section 4.2.2 and 3.6). By focusing on Amplats' high school interventions it is also easier to correlate beneficiary outcomes with the results emerging from the NSC examinations, for which more reliable and detailed official data is available (see section 3.5). Some of the Amplats school projects analysed are learner focused, some are educator focused, some involve the provision of learning and teaching study material (LTSM), or a combination of the above. In great part, the education programme of Anglo American is very similar to what other corporate social investors have been doing in South Africa, by targeting selective groups of learners and schools around their areas of operation (see for instance section 3.7.3 and 3.8.8).

Based on documentation and reports retrieved from the company and its service providers, a brief overview is provided of the major FET interventions of Amplats implemented between 2009 and 2012. This list will succinctly explain the features of the projects, the costs and the number of beneficiaries involved, the service provider responsible and the theories of change (see more in Chapter 3) upon which these interventions were designed:

a) **Supplementary Saturday classes**: These have been occurring through the company Radical Maths and Science since 2003 only in the Rustenburg district, but in 2009 these were expanded also to the operation areas of Atok, Polokwane, Mokopane, and Swartklip. After an initial test and selection process, around 400 Grade 10, 11, and 12 learners were selected to be part of the 'A team' who would take additional classes in mathematics, physical sciences, and business and social skills (including English). A total of 40 or so learners were usually grouped in each area of operation and would meet on Saturdays for a half a day, receive coaching, additional study guides, and prizes at the end for best performance. In 2010 the contract for the Saturday

supplementary classes was passed to the NGO Star Schools, which trained 600 grade 12 learners from all the operations in mathematics, science, and English.

b) **Winter enrichment camps**: These were usually special workshops organised during the school holidays for over 900 learners from the various schools around the various Amplats operations. The programme focused on mathematics, physical sciences, English, and geography. Extra classes run by educators from the local area would normally last for 5 days, and were aimed at addressing some of the content gaps faced by learners in the province due to the poor quality of teaching in their home school, and the inability to cover the ambitious syllabus required for the grade 12 examination. The winter schools would provide study guides, materials, and example questions from the exams, as well as pre- and post-assessment of the learners. Such holiday camps were organised in 2011 and 2012 by the Radical Maths & Science usually in 3-5 different venues across the two provinces, while accommodation, catering, and transport for the learners would be organised separately by the various mines.

c) **Educator training**: This consisted of a programme offered by the Radmaste Centre at Wits University involving a two-week workshop during the winter holidays for 60 teachers from 27 schools from across the Anglo Platinum operations. The course focused on content knowledge in the area of mathematics, science, and accounting, so as to provide educators with more confidence in the subjects they were teaching. The workshop was hands on, interactive, and followed the official Department of Education curriculum. It was followed up by another 2-day follow-up session in September. The programme started in 2012 and was officially supposed to extend over the course of three years (up to 2014); however, the initial plan was interrupted and cut short due to the Marikana happenings, the disruptions and later financial crisis in the platinum industry. The programme was intended to supplement the inadequately trained teachers with supplementary in-service training (see section 3.8.5b) from a reputable teacher training tertiary institution.

As discussed in chapter 3 overview of school intervention, what Anglo Platinum would do through its CSI programme is to supplemnt the weak public education system particularly in the communities around the platinum mines, where new cohorts of local engineers and technicians were expected to be raised. Aside from the above direct interventions implemented through private companies, NGOs, and educational institutions, Anglo American has also been supporting the systemic initiatives of the government, for example, by providing over R1 million in mathematics and science equipment for all the

Dinaledi schools of the North West (see more on the Dinaledi initiative in chapter 5). Programme logic and results frameworks are less explicitly illustrated in the project reports, therefore leaving the readers to infer the theory of change, which otherwise is fairly simple and straightforward for each of the above projects (see chapter 3 for discussion on the theory guiding each of these types of interventions).

The challenge, however, remained that different schools received different treatments (i.e., teacher training, winter schools, infrastructure support) by different service providers (i.e., Star Schools, Radical, Radmaste) and in different dosages (i.e., number of learners and teachers participating in the activities). Information from primary and secondary sources about the various education activities was thus gathered and consolidated for roughly 137 schools in Limpopo and North West that received some kind of support from Anglo American over the past 5 years. In the subsequent quasi-experimental evaluation these will be considered the 'Anglo treatment schools', converted into a dummy variable of 'Anglo yes' for the econometric analysis. A detailed list of Anglo Platinum-treated schools with the various 'dosages' of the different interventions was prepared for the subsequent quantitative analysis.

Most of the above interventions were poorly documented, with only basic information available on the implementation of the activities, including information on some of the participants and progress made by them through pre- and post-learning assessments. As the evaluations were produced mostly by the service providers running the projects, assessments tended to be overall positive, subjective, and relying mostly on qualitative methods to judge achievement of activities and outputs.

This South African case study has highlighted again some of the major problems of evaluation in the development community (donors, NGOs, government), discussed in section 2.8. These shortfalls appear to be even more pronounced in the emerging CSI industry (Besharati, 2015), where evaluations of social programmes lack scientific rigour, independence, and objectivity, and contain biases towards reporting mostly successes (Bamberger, 2009; Riddell & Kruse, 1997; Roche & Kelly, 2005). These agency-commissioned evaluations focus on inputs and spending (Fowler, 2013), implementation processes, and analysis of activities, low-level outputs, and less on development outcomes (Dehn et al., 2003; Kruse, 2003; Masud & Yontcheva, 2005; Pitman, Feinstein, & Ingram, 2005), where more complex impact evaluation methods, involving counterfactuals (Gertler et al., 2011; White, 2005) would be required. The following case study, thus, 're-evaluates' the impact of the above interventions by addressing some of the challenges of classical development evaluation discussed at length in chapter 2, by applying a more rigorous set of methods explained in more detail in the next sections.

4.4 Case study design and methodology

4.4.1 Guiding questions

Emerging from the discussions with the main evaluation client (Anglo American), its service providers, the education authorities, and other stakeholders affected by the platinum mines, the following are the questions that guided the pilot study that assessed the impact of Anglo Platinum's education interventions:

- Were the Amplats projects implemented according to plan? Were the beneficiaries and partners satisfied? Were the right stakeholders and groups benefitting from the interventions?
- What impact did the presence of the mine have on schools in the North-West and Limpopo Province?
- Has Amplats education programme been effective in improving learning outcomes and pass rates (particularly mathematics and sciences) in the targeted schools?
- What can be improved in Anglo Platinum's future education investments? What important lessons can the broader Anglo American group and other mining companies learn from the successes and failures of these programmes?

As most interventions had already occurred, the evaluation was ex-post but still endeavoured to use empirical methods of enquiry. The study was the first to pilot the conceptual framework proposed in this research, which included an evaluation package of participatory applied research methods, qualitative field visits, theory-based quasi-experimental impact evaluation, econometric and meta-analytical techniques. The purpose was to inform government and corporate strategic planning and future investments, improve policies, programming and public systems to better address education challenges in South Africa.

4.4.2 Qualitative field work

The evaluation framework utilised in this study, relies heavily on quantitative approaches such as quasi-experimental evaluation and econometric analysis. Nevertheless, the enquiry began with a preliminary qualitative contextual analysis. This was conducted through classical field visits, though these were not as rigorous and systematic as a full-fledged qualitative case study. Nonetheless this step was important to understand the context, identify the various interventions implemented, unpack the theories of change,

analyse the various factors influencing the outcomes, and discover insights and information which would otherwise be difficult to gauge from a purely quantitative analysis and desktop review.

The study began with three months (September-November 2013) of intense field work in Limpopo and North West to the sites and schools around the Anglo Platinum mining operations. This involved direct observation and site visits to a sample of 22 schools that received support from Anglo American in Rustenburg, Moses Kotane East, Thabazimbi, Leboagomo, Mokopane, Sekhukhune, and Polokwane. During the field visits, semi-structured interviews and focus group discussions were held with various stakeholders involved in Amplats education programmes and in other mathematics and science interventions in secondary schools in Limpopo and North West. These would typically consist of informal conversations with various school principals, mathematics and science educators, curriculum advisors, circuit managers and district managers. Conversations with local education authorities and school management were usually guided by open questions. The discussions were relaxed, informal and non-threatening, following academic research protocols, ensuring confidentiality and protection of subjects and respondents, for instance by not utilising recording devices. Aside from the field trips in Limpopo and North West, interviews were held also in Gauteng with national education authorities and managers of services providers implementing mathematics and science programmes in the two provinces.

4.4.3 Data preparation

Considering the complex impact evaluation methods utilised as part of the study, large datasets needed to be put together that combined data gathered by different institutions and from diverse sources. Some of the data that needed to be merged included:

- Project information
- NSC (matric) exam results
- School administrative, social and economic data

The main units of analysis were schools therefore all the datasets needed to be merged together using the official school EMIS number (unique code given by Department of Education to identify schools). This resulted in an initial mega dataset with a total population of 1412 schools in Limpopo and 385 schools in Northwest, with whom to conduct all the subsequent statistical analysis. Each of the 1797 cases of analysis (schools) was accompanied by a set of other key variables, explained below:

a) Interventions received

Programme and project information included the type of activities, the beneficiary schools, the location, time-period, and costs of implementation. This was provided by Anglo American and its service providers, as well as other organisations (governmental, NGOs, and private sector) implementing programmes to improve mathematics and science in secondary schools. This information was either drawn out of official reports or through interviews with relevant officials.

The primary interventions under analysis were the education programmes of Anglo Platinum implemented between 2009 and 2012. However, during this time frame many other interventions were occurring in Limpopo and North West to address the same outcome of improving mathematics and science learning results in secondary schools. Although the scope and resources of the study did not permit an impact evaluation on each and everyone of these programmes, it was important to be at least aware of these other parallel programmes, so as to control for these potentially confounding interventions in the statistical analysis and isolate the effects of the Anglo Platinum programme from the effects of the other interventions affecting the same outcome (mathematics and science learning). Through the field work in the two provinces and interviews with schools and education authorities, the research team was able to identify some of the major interventions implemented to improve mathematics and science results in secondary schools of the two provinces.

Through literature review and interaction with schools and education authorities in the two provinces, more than 20 confounding interventions were discovered. The concern is that these initiatives by various institutions might have also had an effect on the mathematics and science results of the schools in the region; therefore, the study tried to control for these during the statistical analysis and impact evaluation of the Amplats programme. Some of the bigger and more interesting interventions, would require their own separate impact evaluation to be conducted in the future with additional time, resources, and data.

If data was made available for the external interventions, especially with the specific schools affected, these interventions would be turned into binary dummy variables. A further dummy variable (ANY_intervention_Yes) was created indicating if the school received any intervention by any institution between 2008 and 2012. In the final dataset the variable columns presented in Table 7 were created for every school included in the study:

Table 7. *Intervention Variables used in the Study*

VARIABLE NAME	TYPE	EXPLANATION
Anglo_ScienceLab	Dummy	If the school received science lab and equipment from Amplats
Anglo_Computers	Dummy	If the school received computers and ICT lab from Amplats
Anglo_Other_infrast	Dummy	If the school received any other infrastructure support from Amplats
Anglo_Star_Schools	Scale	Number of learners which participated in the Anglo/Star School Saturday schools in 2011
Anglo_Radmaste_TT	Scale	Number of educators which participated in the Anglo/Radmaste teacher training
Anglo_Radical_2008	Scale	Number of learners which participated in the Anglo/Radical Maths & Science A-Team programme in 2008
Anglo_Radical_2009	Scale	Number of learners which participated in the Anglo/Radical Maths & Science A-Team programme in 2009
Anglo_Radical_2010	Scale	Number of learners which participated in the Anglo/Radical Maths & Science A-Team programme in 2010
Anglo_winterschool_2011	Scale	Number of learners which participated in the Anglo/Radical Maths & Science winter revision camp in 2011
Anglo_winterschool_2012	Scale	Number of learners which participated in the Anglo/Radical Maths & Science winter revision camp in 2011
AngloYes	Dummy	Dummy variable indicating that the school received one of the above or any type of assistance and support from Amplats
Dinaledi	Dummy	The school is a Dinaledi School
Mastec	Scale	Number of educators which participated in Mastec teacher training
RoyalBafokeng	Dummy	School received treatment from Royal Bafokeng Institute
Etc…..	Dummy	School received intervention X…
Any_Intervention_YES	Dummy	Indicating if the school received any intervention by any institution between 2008 and 2012

b) Outcome data: NSC results

As discussed earlier the main outcome indicator used to measure progress of schools was the NSC 12th grade exam results. Thanks to the close collaboration with Umalusi, the researcher was able to gather very detailed and accurate data for all the schools in Limpopo and North West from 2008 to 2012. This included number of learners who wrote the exam, the overall pass rates, and the quality of pass rates in the specific subjects of mathematics and physical sciences. Data on English (first additional language) passes was also observed as this was also found to be correlated to the overall results and performance on the scientific subjects (Botes & Mji, 2010; Taylor & Prinsloo, 2005). Initially data was provided on exact

number of learners who wrote and passed the exams with different marks, but these were later converted into percentages for more appropriate and fair comparisons between different size schools[3]. The scale variables presented in Table 8 were thus computed for all the dataset of schools in North West and Limpopo province.

Table 8. *Outcome Variables used in the Study*

VARIABLE NAME	EXPLANATION
Overall_Passes_2008	% of learners passing the NSC exam in 2008
Bachelor_Passes_2008	% of learners passing the NSC with a pass which allows entry into
Math_Enrol_2008	% of learners registered in the NSC taking mathematics as a subject
Math_Passes_2008	% of learners passing mathematics above 30% mark
Math_Excel_2008	% of learners passing mathematics above 50% mark
Physics_Enrol_2008	% of learners registered in the NSC taking physics as a subject
Physics_Passes_2008	% of learners passing physical sciences above 30% mark
Physics_Excel_2008	% of learners passing physical sciences above 50% mark
English_Passes_2008	% of learners passing English (first additional lang.) above 30% mark
English_Excel_2008	% of learners passing English (first additional lang.) above 50% mark
Overall_Passes_2012	% of learners passing the NSC exam in 2012
Bachelor_Passes_2012	% of learners passing the NSC with a pass which allows entry into
Math_Enrol_2012	% of learners registered in the NSC taking mathematics as a subject
Math_Passes_2012	% of learners passing mathematics above 30% mark
Math_Excel_2012	% of learners passing mathematics above 50% mark
Physics_Enrol_2012	% of learners registered in the NSC taking physics as a subject
Physics_Passes_2012	% of learners passing physical sciences above 30% mark
Physics_Excel_2012	% of learners passing physical sciences above 50% mark
English_Passes_2012	% of learners passing English (first additional lang) above 30% mark
English_Excel_2012	% of learners passing English (first additional lang) above 50% mark

[3] The analysis of NSC enrolments and passes used in this study is based on percentage rates, however it needs to be acknowledged that if absolute numbers of learners from each school would be used as the measure, potentially different results could also emerge from this study.

c) School administrative and demographic data

In order to undertake some of the econometric and matching operations, and have enough control variables (covariates) for the impact evaluation (see more in next sections), it was important to gather as much information about each of the schools in the dataset, including demographic, social, economic data, resource availability, characteristics of learners and educators, management, and other aspects of the schools. This was gathered primarily through the EMIS of the National Departments of Education. Some data was sourced also through StatsSA and the Provincial EMIS offices. The EMIS data used was taken from the ASS and SNAP surveys (see section 3.8). Since the information varied over the years, the EMIS data from 2010 was utilised as a mid-term reference point for when interventions were implemented.

Caution was taken with the EMIS data, as this was not known to be always accurate and reliable, as schools sometime under-reported or over-reported, because education funding and resource allocation is often linked to these school characteristics (Bisgard, 19 September 2013, personal conversation; Woolman & Fleisch, 2006). If the data was not accurate or useful, such variables were often taken out of the analysis. In many instances data needed to be cross-referenced, double-checked, and re-calculated, transforming and re-coding some of the variables during the data cleaning process. Whenever possible absolute numbers were converted into percentage (i.e., total number of learners or teachers) in order to have a more useful variable to compare schools of different sizes and also to make the results easier to understand for the common reader. A mega dataset was finally generated with the variables (Table 9) recorded for each of the nearly 1800 schools in the population of secondary schools in North West and Limpopo used in the study.

Table 9. Explanatory Variables Used in the Study

NAME	TYPE	EXPLANATION & NOTES
NatEMIS	Nominal	Unique school ID number used to linked datasets together
School_name	Nominal	Name of institution
Province	Nominal	Province where School is situated
District	Nominal	District Municipality where School is situated
Circuit	Nominal	Circuit where School is situated. This changed over the years
GIS_latitude	Scale	Latitude of school GIS coordinate
GIS_longitude	Scale	Longitude of school GIS coordinate
Quintile	Ordinal	Coded as 1, 2, 3, 4 and 5 reflecting the social-economic status of the school with 1 being the poorest and 5 the richest.

Table 9. Explanatory Variables Used in the Study (continued)

NAME	TYPE	EXPLANATION & NOTES
Section_21	Dummy	Indicates a type of advanced status and autonomy of school. This variable was eventually taken out as the original data indicated strangely that all Limpopo schools are section 21, therefore not useful variable and probably suffering from transcription errors.
Urban_Rural	Dummy	Indicates the location of a school in relationship to urban or rural areas. This variable was eventually taken out as most of the data was missing or incomplete.
FeeORnot	Dummy	Indicating if the school is a 'no fee schools'
AverageFees	Scale	Average yearly fees charged by school for grade 10, 11 and 12
SchoolType	Dummy	Indicating if the school is public or private
Specialisation	Ordinal	Coded as 6 = Technical 5 = Commercial 4 = Comprehensive 3 = Ordinary 2 = Agriculture 1 = Arts
Land_ownership	Dummy	Indicating if the schools is on privately owned land
Building_ownership	Dummy	Indicating if the school is on privately owned building
Bantustan_Yes	Dummy	Indicating if schools used to fall under previous homeland administration.
Total_Learners	Scale	Total number of learners in grade 10, 11 and 12
EngAfr_HLang_learners	Scale	% of learners whose home language is English or Afrikaans
Per_black_learners	Scale	% of learners who are black and other HDSA racial groups
Av_Learners_Disable	Scale	Average number of learners with academic disabilities in grade 10, 11 and 12
Av_Pregnancies	Scale	Average number of learners who fell pregnant in grade 10, 11 and 12. Eventually variable dropped out of the analysis as considered not useful
Learner_repeaters	Scale	Number of learners repeating grade 10, 11 and 12. Eventually variable dropped out of the analysis as considered not useful.
Learner_dropouts	Scale	Number of learners dropping out of grade 10, 11 and 12. Eventually variable dropped out of the analysis as considered not useful.
Learner-Educator_ratio	Scale	Number of learners divided by number of educators
Total_MST_Educators	Scale	Number of educators dedicated to maths, science and technology
Percent_MST_Educators	Scale	% of educators dedicated to the teaching of maths, science and technology from the total number of educators in the school
Learner-MSTedu_ratio	Scale	Ratio of maths, science and technology educators available to total number of learners in the school
PercAcDegree	Scale	% of educators with a university degree

Table 9. Explanatory Variables Used in the Study (continued)

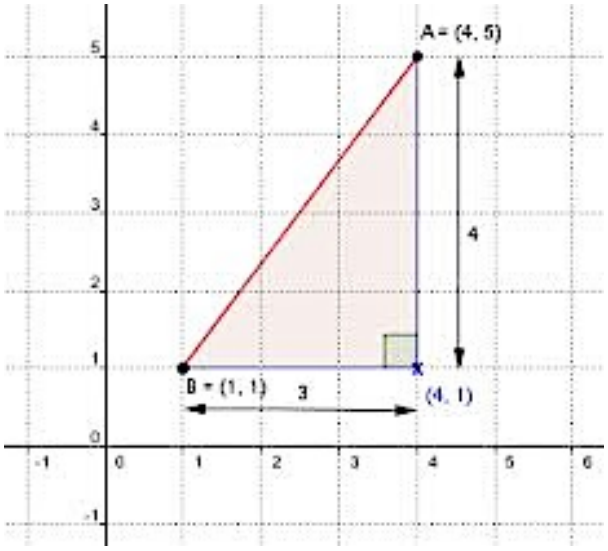
NAME	TYPE	EXPLANATION & NOTES
Edu_exper_<15	Scale	Number of educators with less than 15 years of work experience
Edu_exper_>15	Scale	Number of educators with more than 15 years of work experience
Perc_Young_Edu	Scale	% of educators with less than 15 years of work experience from total number of educators
Perc_Old_Edu	Scale	% of educators with more than 15 years of work experience from total number of educators
Total_Foreign_MSTedu	Scale	Total number of foreign (ie. Zimbabwean) educators teaching Maths, Science and Technology subject
Per_Foreign_MSTedu	Scale	% foreign (ie. Zimbabwean) educators teaching Maths, Science and Technology subject
Foreign_Educ_Yes	Dummy	Indicating if the school uses foreign (ie. Zimbabwean) educators
Science_Lab	Dummy	If a science lab is present in the school
Computer_Lab	Dummy	If an ICT lab is present in the school
Media_Library	Dummy	If a media library is present in the school
Multi-group_class	Dummy	Schools in which more than one class group needs to share one classroom
SGB_approveCurriculum	Dummy	School with curriculum approved by SGB
Time_tables	Dummy	School with Time Tables available
Curriculum_Plan	Dummy	School with Curriculum Plan available
Curriculum_Test	Dummy	School with regular learning assessments available
Curriculum_Monitoring	Dummy	School with curriculum coverage monitoring ongoing
qualityofMaths	Dummy	School with access to quality Maths programme
qualityofScience	Dummy	School with access to quality Science programme
Integrate_ICT	Dummy	Integration of ICT facilities in teaching process
Teaching_aids	Dummy	Availability of teaching aids for educators
Distance_Mine*	Scale	Calculated using the GIS coordinates of the schools and the platinum mines using haversin formula. See more below for details.
50km_from_Mine*	Dummy	Dummy variable indicating if school is within 50km from a platinum mine. Calculated using the distance from mine variable above.
District_Location*	Dummy	Each one of the district municipalities of North-West and Limpopo was converted into a dummy variables namely: Greaterdelareyville, Greatertaung, Kagisanomolopo, Kgetlengriver, Lebowakgomo, Lichtenburg, Madibeng, Mafikeng, Maquassihills, Matlosana, Mogalakwena, Mopani, Moretele, Moseskotaneeast, Capricorn, Potchefstroom, Polokwane, Sekhukhune, Rustnburg, Taledi, Vhembe, Waterberg, Zeerust. See more below.

From the field research it was evident that the location in which the school was situated could significantly influence the learning results, as each area received attention and support by different external institutions, but also each circuit and district education office had different officials with different levels of commitment and enthusiasm for the initiatives, which could affect positively or negatively the outcome of the intervention. As the urban/rural data did not prove to be complete and accurate, the other option was to convert the categorical variables of the locations into numerical values. As the circuits were far too many and the provinces were only two, dummy variables were created for all the 23 district municipalities in Limpopo and North West. Each district contained anything between 10 schools (Kagisano Malopo-NW) and 233 schools (Capricorn-LP). This exercise was also useful to corroborate the South African literature (Fleisch, 2006; Taylor et al., 2003; Taylor & Prinsloo, 2005) on district development, discussed earlier in section 3.8.7.

Another geographic factor found to have an impact on the schools (see later sections on findings) was the proximity to the mining operations. In order to calculate the distance (km) between the schools and the mines, and subsequently construct the dummy variables '50 km from mines', the GIS coordinates (Lat/Lon) of each school and each platinum mine was gathered and a basic formula of trigonometry was applied. Assuming a flat surface a straightforward Pythagoras theorem could be used to calculate the distance (d) between the mines (m) and the schools (s) as in [12]:

$$d = \sqrt{(Lat_m - Lat_s)^2 + (Lon_m - Lon_s)^2} \quad (12)$$

Figure 34. Geometrical representation of Pythagoras formula.



However the surface of the earth is spherical and when calculating distances above 17km, the Pythagoras formula (Figure 34) will become inaccurate and cause errors. Therefore, the Harvesin formula was used to calculate with more precision the distance between the schools and platinum mines in North West and Limpopo. The Harvesin formula is as follows [13]:

$$d = 2r \arcsin\left(\sqrt{\operatorname{haversin}(\phi_2 - \phi_1) + \cos(\phi_1)\cos(\phi_2)\operatorname{haversin}(\lambda_2 - \lambda_1)}\right), \quad (13)$$

where

$$\operatorname{haversin}(\theta) = \sin^2\left(\frac{\theta}{2}\right) = \frac{1 - \cos(\theta)}{2}$$

- *d* is the distance between the two points (along a great circle of the sphere),
- *r* is the radius of the sphere,
- ϕ_1, ϕ_2: latitude of point 1 and latitude of point 2
- λ_1, λ_2: longitude of point 1 and longitude of point 2

d) Final data cleaning and rendering

To conduct econometric analysis on the data gathered from the various institutions, most of the variables needed to be converted into numerical values. Some were turned into binary dummy variables, some into scale variables (absolute numbers and percentages), and some into ordinal variables (such as school quintile and school specialisation).

In order to utilise some of the more sophisticated statistical techniques such as PSM, it was important that the dataset did not have any missing or incomplete values. In the original datasets provided by the various institutions, there was a lot of missing data, so the researcher had to take decisions on how to treat the missing data. For many of the variables such as the information about infrastructure, resources, curriculum, language, and social-economic status of schools and learners, missing values were estimated by using the mode (majority value) of the cases, the mean (average), or the value which appeared most logical based on theory and field work. In the event where there were too many cases of missing data which could not be estimated and the variable did not seem to have a major impact on

results (such as with the variables of the former Bantustan schools, the percentage of educators with academic degrees, the urban/rural location, repeaters and drop-outs), the variable was omitted from the matching process. If the missing and problematic values were only few (such as with some of the NSC results), the specific cases would be deleted from the dataset, as long as these did not affect the schools receiving interventions (which were most important for the evaluation).

Finally, the dataset was checked for any outliers and anomalous cases and corrections were made accordingly. By the end of the data cleaning process, the total numbers of secondary schools from Limpopo and North West included in the analysis was brought down from 1796 to 1536; and the total number of Anglo Platinum-supported schools was reduced from 137 to 134. Both the treatment sample as well as the population sample, from which the control groups were extracted, were large enough to arrive at statistically significant results during the next stages of the impact evaluation.

4.4.4 Impact evaluation methods and techniques

Once the dataset for the complete population of schools in North West and Limpopo was prepared and cleaned, the required statistical analysis could be conducted in order to measure the impact of the Amplats interventions. As the evaluation taking place was ex-post and a randomised experiment was never planned at the start of the programmes, the study relied on quasi-experimental methods, which were easier, cheaper, and more practical to implement (Baker, 2000; Sanson-Fisher, et al., 2007), but nevertheless required access to a lot of administrative data (see also section 2.13), which luckily was the case in this situation. The use of un-obstructive ex-post analytical methods also prevented the study from unwarranted *Hawthorne effect* (Stand, 2000) and *John Henry effect* (Saretsky, 1972)[4], which could have occurred in the control and programme group if a real-life experiment was set up. Based on the evaluation framework and the set of methodological approaches discussed in chapter 2, the Amplats case study relied on quantitative techniques such as multivariate regression models, a geographically adapted cut-off point design, PSM, DID, student *t*-test, and approaches for effect size calculation and interpretation borrowed from meta-analysis. The theory behind these methods is discussed in detail in chapter 2 of the research; however, the specific application of the framework to this particular case study will be illustrated in the next sections on the findings of the Amplats evaluation.

[4] See more in page 54

4.5 Evaluation findings

4.5.1 Results from the qualitative field work

This initial section highlights some of the findings from the fieldwork. It represents the synthesis of the views and inputs gathered from the interaction with the local stakeholders, through interviews and focus group discussions. For the purpose of this PhD dissertation only a segment of these findings are reported in this chapter, as much of the feedback was related to project delivery, thus useful to Anglo American and its partners, but less relevant to the academic reader. As such only the qualitative findings that are pertinent to understanding the broader theories discovered and to explaining some of the results of the quantitative findings will be captured in this chapter. For more details about the implementation of the programme and further feedback from project stakeholders, the reader is invited to read the more in-depth evaluation report that was prepared for Anglo American and the broader public (Besharati, 2014).

Some of the following qualitative information needs to be considered within the bounds of subjectivity as they represent the opinions of the individuals with whom the researcher engaged during his field visits. Nonetheless these interactions and inputs were important and useful to explain more vividly the background and local context for the evaluation, and shed more understanding and insights on the quantitative enquiry, validating some of the theories that emerged from the results of the regression and experimental work that occurred subsequently in the study.

a) Selection issues

One of the key elements is the way in which the beneficiaries of the Anglo American Platinum education programme were selected and chosen. This process had an inevitable effect on the overall impact of the interventions, as well as on how the evaluation of the programme is now being conducted (see section 2.13). Due to limited resources, the Anglo Platinum programme could not accommodate all the communities in Limpopo and North West, and therefore a selection process had to be undertaken in one way or another.

The first level of selection that occurred was at the school level. According to Anglo Platinum's official CED policy, the first communities to receive support from the company are the ones located within a 50 km radius of the mine. This implies that certain areas (Rustenburg, Mokopane, Thabazimbi, Sekhukune, etc.) and certain tribal groups (i.e., Bafokeng, Bengwenyama, Langa Mapela) that are situated closest to the mines benefit the most from the mining companies' social investments. This policy

is obviously criticised by those schools and communities located farther away from the mine. Whatever the fairness of this policy, it is true that those communities and schools closest to the mines are also the ones that suffer the most from the negative externalities of mining operations.

The next level of selection occurred at the level of individual participants from the short-listed schools. Aside from a few teacher-training interventions, the majority of education programmes focused on learners. This presented a problem, as resources were limited and the service providers could only accommodate a small group of learners in their programmes. In the words of one school principal, the interventions were "just a drop in a big ocean" and therefore played a marginal role in affecting overall pass rates within the school system. Once the selected learners passed grade 12, there would be more learners with the same problems, and everyone would be back to square one. Education stakeholders in the two provinces have stated that it would be better to assist all of the learners and not just the "lucky few".

However, this utilitarian argument can be contested, as others will argue that once the intervention is diffused within the broader population it becomes less effective and more difficult to perceive its impact. From a practical perspective, one can also argue that in a situation with limited funds it is best to focus on fewer promising individuals who can reap the most benefits from the intervention and provide good returns on investment for the company's human development and future business operations.

The selection of learners for participation in various programmes can thus be a complex and sensitive undertaking for any donor or service provider, with many ethical and political considerations to take into account. For this reason, the selection process is usually delegated to the education authorities, such as school principals and circuit managers (with some general guidelines from the corporate donors), or done by the service providers (i.e., Radical Maths & Science), which often conduct entry tests and assessments.

Regardless of the methods employed in the selection process, usually the brightest and best performing learners (the crème) in each class and school are selected to take part in the programme. This phenomenon is not limited to Anglo American: almost all corporate institutions and NGOs running education interventions in the provinces function in this mode.

This causes two major problems. The first is that programmes can be accused of being discriminatory: the learners who need help the most are left behind, while the learners who would have performed well anyway receive even more support. There is a strong correlation between interventions and the increase in the number of *bachelor passes*, but less correlation between interventions and the *overall pass rates* of schools.

The second problem was one around methodological validity and bias which hinges on the evaluation design. All the service providers contracted by Anglo American Platinum (Radmaste, Radical Maths & Science, Star Schools) diligently conducted pre- and post-tests, continually demonstrating the success and positive growth of the beneficiaries as a result of their participation in the programmes. From a scientific point of view, however, these *before and after* assessments must be taken with a pinch of salt. They are done without a counterfactual (control group) and therefore claims of causality need to be considered with caution, as those participants would probably have improved their results even in the absence of the programmes. From a methodological point of view, the above process will inevitably fall victim to *selection bias* and thus should be treated differently during the impact evaluation work.

These selection parameters for schools and learners have not been applied strictly across the two provinces. During the field visits it became evident that some of the more active district and circuit managers and school principals advocated for the inclusion of all the learners in their schools even if they were outside the 50 km radius. Some of these schools have also been resourceful and found ways to multiply and cascade the support that a few learners received from the mining companies by making photocopies of the materials and asking the learners who participated in workshops to share their knowledge with those of their peers who were not able to participate, for example, in the winter schools.

b) The key role of teachers and school managers

Although the focus of the study was on Amplats and other external interventions, what emerged very strongly in the discussions with local stakeholders, was that ultimately the single most important factor determining better learning results was *teachers*. This is consistent with the debates in the overall education literature (Hanushek, 2010; Kremer et al., 2013; Simkins, 2010; Taylor, 2011; World Bank, 2011) and as well as the discussion presented in section 3.8.5 of this research. The quantity, quality, and commitment of educators in a school is the prime factor driving pass rates up, and therefore this study took the liberty to investigate this aspect a bit further in the field work.

The shortage of mathematics and science educators in the system, underlines the importance of recruitment and training programmes such as the 'Funza Lushaka' scholarship programme (see section 3.8.5a). From the field work undertaken in both Limpopo and North West, it was also visible that there was a large presence of Zimbabwean and other foreign teachers in the schools to compensate for the lack of well-qualified South African teachers in the system (CDE, 2010; Spaull, 2013). The issue of Zimbabwean teachers is highly sensitive, political, and controversial, as it impacts also on South Africa's immigration policy. On the one hand some education officials would like to curb employment of foreign teachers in order to give more of the much needed jobs to local South Africans, and on the other hand

many school principals have shown that they prefer to go through the extensive bureaucratic procedure to hire foreign educators, who appear to perform better than South African educators. This sensitive policy debate might be the subject of future academic research.

When assessing the quality of the teachers many differing opinions surfaced in the interviews with the school managers, such as the preference between young and enthusiastic versus old and experienced educators. Another major issue of debate during the field conversations was whether the new educators trained in the highly-theoretical and content-rich universities are better than the old educators trained pre-1996 in the teacher-training colleges, which were allegedly stronger in pedagogical methods and practical experience? This is also consistent with some of the previous academic debates (Chisholm, 2009; Pelzer et al., 2005; Sayed, 2004; Wolhuter, 2006), discussed in section 3.8.5b.

Although not much drastic change could occur in the current cadre of educators of the country, what is perhaps more relevant for the government and private investors in South Africa is to investigate what are the best in-service teacher training and upgrading programmes available (Ono & Ferreira, 2010; Parker, 2002; Wolhuter, 2006). Currently in South Africa there are many different educator development programmes, models, and approaches. Only in Limpopo and North-West the case study identified numerous initiatives run by different institutions such as those of Radmaste (Amplats), Royal Bafokeng Institute, Mastec, Dinaledi, Mpower (Kumba/TCF), to just name a few. Considering the centrality of educators in the schooling system, more in-depth evaluation of different types of teacher development programmes would be useful in order to identify which model seems to have more long-term impact on learning results (see also Taylor, 2011).

Strongly related to the quality and performance of educators was the issue of leadership and commitment of school principals. Highlighted also by the academic literature (Christie et al., 2007; Leithwood et al., 2004; Taylor et al., 2003; Taylor et al., 2013) and by the diagnostic report of the National Planning Commission (2011), it was very visible even during the field work that the best-performing schools were usually the ones with a strong, dedicated, and competent school director. This was even perceptible by the simple appearance of the school, the cleanliness, order, and neatness of the facilities. As discussed also by Taylor et al. (2012), strong school organisation and management, solid system of accountability of teachers, regular learner assessments, feedback and engagement with parents, organisation of extra-curricular activities, fund-raisers, additional workshops, and other special initiatives to improve learning outcomes, were clearly factors which emerged out of the most successful schools visited.

Local stakeholders suggested various interventions such as prizes and awards for best principals, teachers, and learners that could potentially function as incentives for improvement of educational outcomes. The importance of school management, good governance, and leadership in the achievement

of high learning results is consistent also with much of the rest of the South African education literature (Christie et al., 2007; Simpkins, 2010; Taylor, 2011) discussed in chapter 3.

c) Negative externalities from proximity to mines

As the research team immersed itself in the field and interacted with various stakeholders, a very interesting theory emerged, especially from conversations with school principals. Rather than acting as a stimulus to better pass rates, the presence of a mine close to a school actually has a negative effect on education outcomes. There are some obvious negative consequences to having a mine in a community, such as the frequent protests and unrest that disrupt local community affairs, including schooling activities.

But a more interesting phenomenon is the fact that when a mine is present in a rural community, people develop high expectations for employment and the improvement of livelihoods through the linkage with that mine. Families also know that the jobs required by the mine are more technical and engineering related. As a result, there is a strong push for learners in mining areas to take subjects such as mathematics, physical and natural sciences, geography and technology, which would allow them to later enter university courses and apprenticeship programmes that would increase their chances of employment in the lucrative mining industry.

This tendency, however, also sees learners who do not have an aptitude for mathematics and science enrolling in scientific and technical subjects in the NSC examination, leading to the overcrowding of learners in subject areas that are not properly resourced with good educators, facilities, and materials. Since more ill-prepared learners take these subjects, more learners around the mines actually end up failing the grade 12 examination compared to other areas.

This is consistent also with the general trend observed earlier in section 4.2.2 at provincial level that an increase in pass rates is strongly linked to a decrease in NSC exam enrolments for mathematics and science. The researcher felt obliged to further test this theory through the quantitative analysis laid down in the next few sections.

4.5.2 Results from the econometric analysis

As discussed in section 2.10 of this research, econometric non-experimental approaches are often used to measure impact of interventions. However, these are usually not very precise (Gertler et al., 2011), as

they are based only on observable variables. In such cases, one can never fully control for unknown extraneous variables, which may have significant confounding effects on the outcome (Todd, 2012). Therefore, experimental and quasi-experimental approaches are preferred over regression approaches for the purpose of calculating net impact. Nevertheless, running a regression with the variables available is a useful first step to have some idea of the factors that could potentially be more closely associated with the outcome (y) variable and how significant these correlations are. This procedure is similar to other education production functions conducted by other South African education analysts discussed in chapter 3 (i.e., van der Berg, 2008; Gustaffsson, 2007; Taylor, 2013), where relationships between different learner, teacher, school, community characteristics are observed in relationship to learner achievement.

Thus, before the more rigorous experimental evaluation was conducted on the Amplats programme, a simple linear step-wise regression was run using the list of independent variables gathered in the study which included data on the various interventions implemented in Limpopo and North West as well as all the school administrative, social-economic, and demographic data gathered from the education institutions (see section 4.4.3).

The main purpose of running the regressions was to have a rough indication on the correlation of the different variables on the learning outcomes and pass rates in Limpopo and North West schools. Of particular interest were the beta-coefficients and the significance levels of their respective probability values (p-value). Four regression models were run using the following variables as the dependent (y) variables:

- NSC mathematics passes (above 30%)
- NSC mathematics excel (above 50%)
- NSC physical science passes (above 30%)
- NSC physical science excel (above 50%)

Almost all the interventions, school administrative, demographic factors (a total of over 130 explanatory variables for over 1400 schools) were used and tested in the regression model. The objective was to have a rough indication of which factors were significantly and strongly correlated to learning results, and thus explore these aspects further during the subsequent experimental work. The important variables would also be used as covariates for the PSM and as controls during the impact assessment work (see next section).

Care was given to duplication and choosing variables which were not too similar to other variables and which could present issues of multicollinearity. Through SPSS a step-wise regression model was used using a 95% confidence level, on all four outcome variables, and usually by the seventh or eighth step, most of the variables would be dropped and the only ones which would remain were the

ones which were statistically significant in the model. Below are the regression tables which indicate also the coefficients of the variables that are statistically significant (Tables 10-13), thus giving a good indication of the factors which have the most impact on learning outcomes. More interpretation and explanation of the results is presented the bottom of the tables.

Table 10. *Econometric Analysis on the Impact of Various Variables on Maths Results: Regression on Maths Passes (above 30% score)*

Model		Unstandardised Coefficients		Standardised Coefficients	t	Sig.
		B	Std. Error	Beta		
8	(Constant)	.401	.030		13.561	.000
	English_Excel_2012	.377	.028	.330	13.412	.000
	AverageFees	2.196E-005	.000	.164	6.207	.000
	Distance_mine	.000	.000	.088	3.266	.001
	LearnerEducator_ratio	-2.504	.492	-.121	-5.086	.000
	Math_Enrol_2012	-.128	.028	-.108	-4.503	.000
	EngAfr_HLang_learners	.001	.000	.094	3.716	.000
	OwnerLand	.160	.050	.076	3.174	.002
	VHEMBE	.052	.019	.075	2.794	.005

Table 11. *Econometric Analysis on the Impact of Various Variables on Maths Results: Regression on Maths Excel (above 50% score)*

Model		Unstandardised Coefficients		Standardised Coefficients	t	Sig.
		B	Std. Error	Beta		
7	(Constant)	.276	.040		6.931	.000
	English_Excel_2012	.167	.018	.228	9.177	.000
	AverageFees	1.676E-005	.000	.196	7.578	.000
	Per_black_learners	-.002	.000	-.165	-6.168	.000
	LearnerEducator_ratio	-1.467	.310	-.111	-4.728	.000
	Quintile	.013	.004	.073	2.857	.004
	VHEMBE	.032	.010	.071	3.050	.002
	TotalMSTEduc	.005	.002	.068	2.916	.004

Table 12. *Econometric Analysis on the Impact of Various Variables on Physics Results: Regression on Physics passes (above 30% score)*

Model		Unstandardised Coefficients		Standardised Coefficients	t	Sig.
		B	Std. Error	Beta		
7	(Constant)	.479	.028		17.276	.000
	English_Excel_2012	.346	.029	.288	12.029	.000
	VHEMBE	.153	.017	.211	9.102	.000
	AverageFees	1.787E-005	.000	.127	4.813	.000
	LearnerEducator_ratio	-3.407	.517	-.157	-6.589	.000
	WATERBERG	-.143	.039	-.084	-3.623	.000
	OwnerLand	.199	.053	.090	3.754	.000
	EngAfr_HLang_learners	.001	.000	.084	3.294	.001

Table 13. *Econometric Analysis on the Impact of Various Variables on Physics Results: Regression on Physics excel (above 50% score)*

Model		Unstandardised Coefficients		Standardised Coefficients	t	Sig.
		B	Std. Error	Beta		
8	(Constant)	.281	.049		5.800	.000
	English_Excel_2012	.188	.019	.240	10.065	.000
	AverageFees	1.942E-005	.000	.212	8.127	.000
	LearnerEducator_ratio	-2.236	.323	-.158	-6.924	.000
	Per_black_learners	-.002	.000	-.118	-3.486	.001
	VHEMBE	.064	.010	.136	6.158	.000
	Dinaledi	.082	.017	.108	4.775	.000
	EngAfr_HLang_learners	.001	.000	.098	3.049	.002
	OwnerLand	.094	.033	.065	2.844	.005

The econometric analysis above confirmed theories and results that the international and South African education literature had previously shown (see chapter 3), but it also revealed some interesting new findings.

The most influential factor influencing both mathematics and science pass rates is the pass rates registered by the same schools in English language. This confirms the various theories around the importance of strengthening the command of language for all other learning to take place (Cummins, 2009; Myburgh, Poggenpoel, Taylor & Coetzee, 2013, Van Rensburg, 2004). This is further confirmed by another regression variable strongly correlated to pass rates namely the percentage of learners whose home language is either English or Afrikaans.

As previously explored by other South African academics (Shepherd, 2001; Simkins, 2010; Taylor, 2011; Van der Berg, 2008), these four regressions demonstrate that social-economic factors are among the most influential variables on learning results. At the top of the list is the average school fees, which show that the more families pay (and thus resource, incentivise, and bring to account the schools) the higher the pass rates are in those schools. Percentage of black learners is negatively correlated to mathematics and science results, thus underlying the still vast racial-social divide which affects the economy of the country. Furthermore, the higher the school quintile the better the pass rates becomes.

Also widely discussed in the literature (Deacon & Simkins, 2010; Chisholm et al., 2005; Gustafsson, 2007; Spaull, 2011; Taylor, 2009, 2013), is the critical role of teachers in the learning process. Similar to Taylor (2013) education production function, the regression models show that there is a negative correlation between an increase of class size / educator-learner ratios and a decrease in the learning results. A large presence of mathematics, science, and technology teachers in the school (Total MST_Edu) seem to also have a small effect on quality mathematics pass rates.

Interesting enough, none of the many external interventions in Limpopo and North West, including the substantial investments by Anglo Platinum, had a significant impact on learning results of schools. The only intervention which seem to have some effect on science passes is the Government's Dinaledi programme, which confirms the findings of the World Bank evaluation (Blum et al., 2010) and CDE studies (Simkins, 2010), and warrants probably a further update and re-visiting into the long-term impact of this initiative, considering also the programmatic evolution and new delivery approaches that the programme undertook since 2008.

Private ownership of the land seems to also have a small effect on pass rates. Section 14 of the South African Schools Act (1996) makes provision for public schools that are situated on private property. These can be farms, nature reserves, land belonging to tribal authorities (i.e., the Royal Bafokeng). But in great part section 14 schools tend to be the former Catholic schools and the schools managed by churches and religious groups. These seem to have better learning outcomes than the rest of

the public school system, possibly because of the commitment and values that the staff and learners might have to the education process. Further research in this area could be highly interesting to explore.

Of all the 23 district municipalities which were included as dummy variables in the regressions, the only one which seems to have a strong correlation to pass rates is the district of Vhembe, which is the district further most North in South Africa, close to the border of Zimbabwe. Ironically this rural and poor area of Limpopo, formerly part of the Bantustan state of Venda, had a positive correlation with pass rates, which would be worth investigating further. Can this be related to the education policies of the former Venda homeland[5] or can it be because of the district's proximity to Zimbabwe, thus access to better trained mathematics and science educators?

Vhembe district is also far from the mining operations therefore this would exclude a positive impact stemming from the education programmes of Anglo Platinum and other mining houses. Interesting enough these regression models show that pass rates increase the farther away the school is from the platinum mine. This goes to further strengthen the theory discussed in the qualitative findings on the negative impact of the mines in relation to school results (see section 4.5.1c). Also as explored by the descriptive statistics and the provincial trends presented at the beginning of the chapter, as well as the regression analysis, an increase in mathematics enrolments (possibly due to the presence of mines in the area) is strongly correlated also with a decrease in the pass rates. This will be further explored in the next section.

4.5.3 Results from the quasi-experimental evaluation

a) Problems with cut-off point experiment

As discussed in chapter 2, in the absence of a RCT, one of the more reliable quasi-experimental methods, is that of RDD (Dehejia, 2013; Hombrados & Waddington, 2012; Shadish & Cook, 2009). This method works however if there is a very strict scale-based cut-off point which determines a quasi-random participation and non-participation in the programme (see section 2.13b).

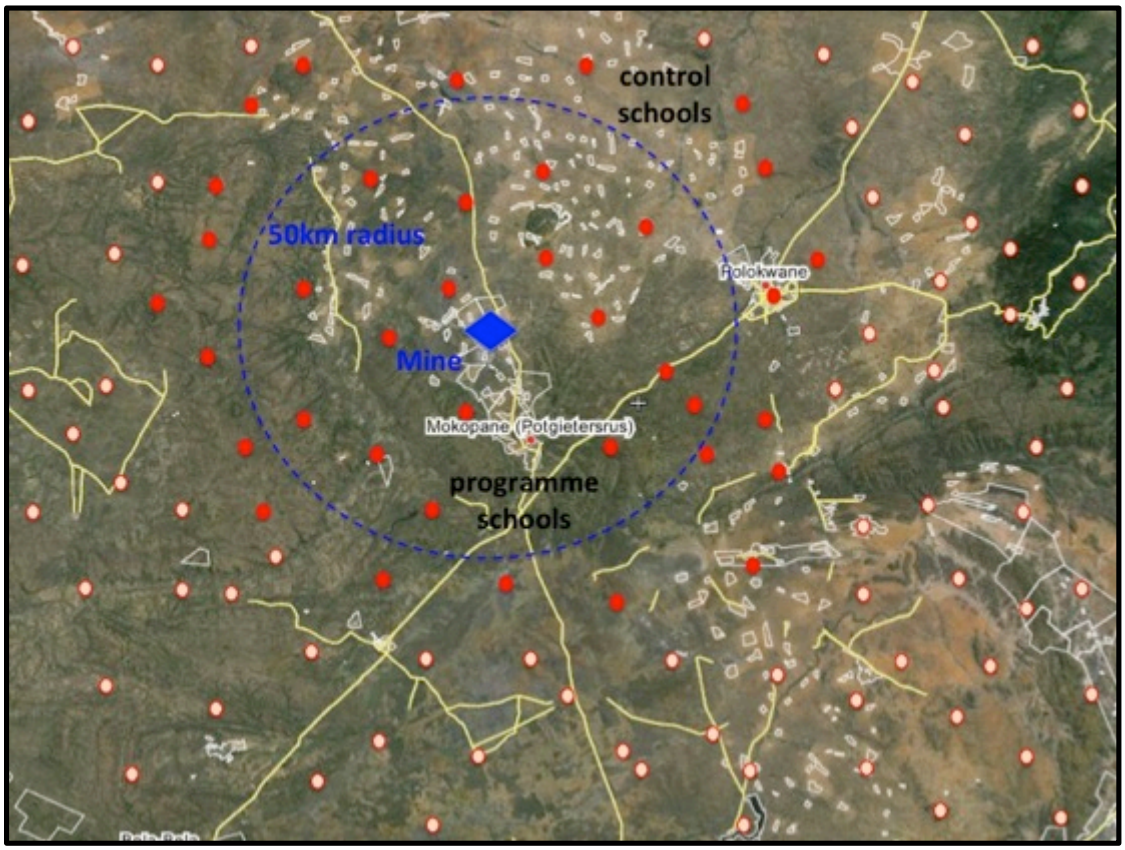
Considering the particular nature of the study involving mines and affected communities, the case study tried to experiment with the same concept of the cut-off point design technique, but applied to

[5] A similar finding also emerged from the World Bank (2010) impact evaluation of the Dinaledi initiative, which showed that schools in former Bantustan areas performed better than schools under former House of Assembly administration.

a continuous scale of geographical distances (Kms). We have seen earlier that one of the school variables calculated using the GIS coordinates was the distance of the schools from the mine.

According to the mining legislations as well as the CED strategy of Anglo Platinum, the communities (and therefore also the schools) that should receive support and compensation from the mine are the ones within 50km radius. If such selection of participants into the Amplats education programme was applied strictly and there was no spillover between programme and control group, we thus could utilise the cut-off point design technique. This would imply an imaginary geographic circle with 50km radius around the mine, where the schools that are part of the intervention and close to the border of the circumference are theoretically very similar to the schools that are just outside the circumference and did not get into the programme. Thus we would have a good experimental sample and control sample to compare against.

Figure 35. Example of application of RDD to geographic scale within mining context.

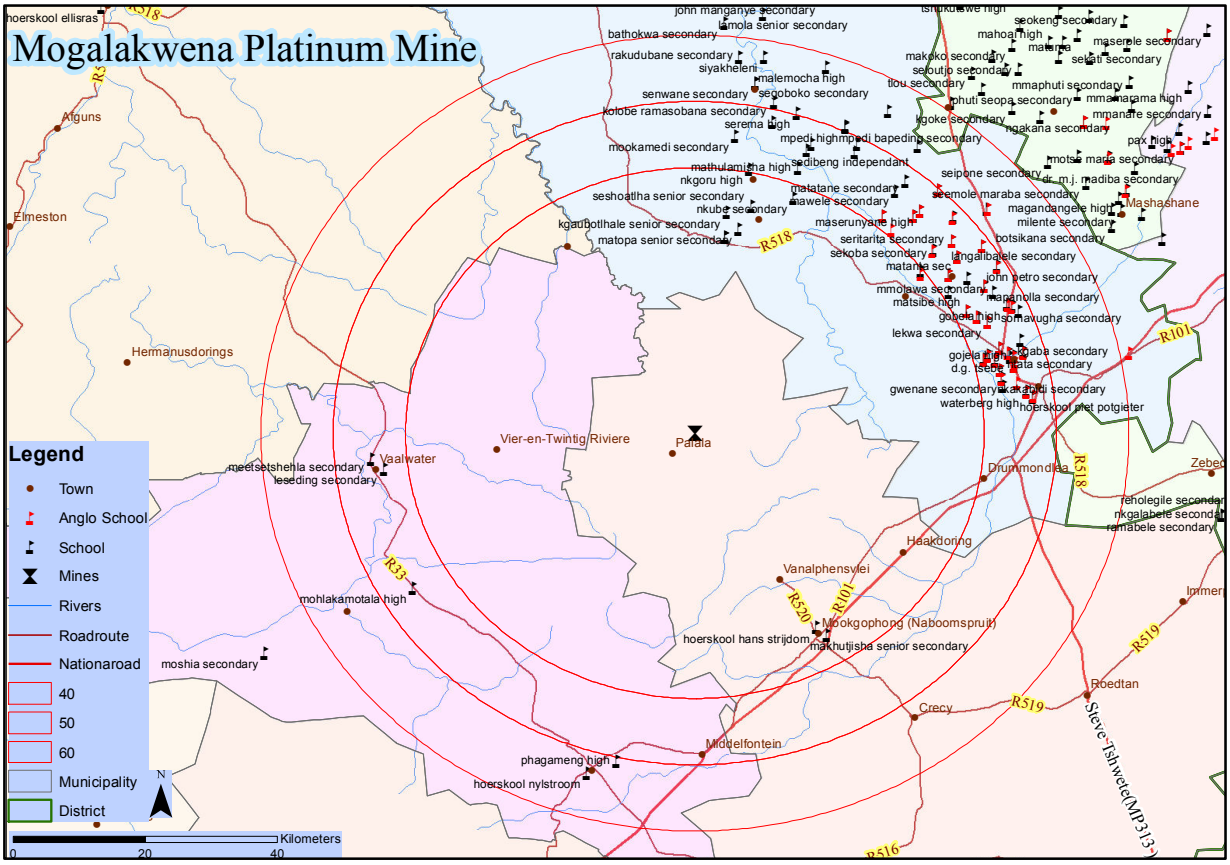


Source: Authors own compilation (2013)

As a first step, the study calculated the GIS co-ordinates of the various platinum mines and the schools both serviced and not serviced by Anglo American Platinum. With the assistance of StatsSA, these mines and schools were plotted on maps of Limpopo and North West to observe whether this 50 km selection criterion was applied in a strict and consistent manner. The schools in red (see Figure 35) represent the 'Anglo schools', versus the 'non-Anglo schools' in black, all placed in three buffer zones of 40 km, 50 km, and 60 km from the mines. These imaginary lines could serve the purpose of allowing for the 40 km to 50 km schools to be treated as the programme group and the schools in the 50 km to 60 km buffer to be used as the control group in a cut-off design evaluation.

Maps for all Anglo Platinum mining and processing operations are available in Besharati (2014) research report; however, Figure 36 is an example of just one of these plotted maps where the geo-locations of the high schools serviced and non-serviced by the company are shown around the Mogalakwena mine.

Figure 36. Maps of geo-locations of Mogalakwena mine vis-à-vis treated and untreated schools.



Source: Statistics SA (2013)

It is apparent from even a glance at these maps that the ’50 km rule’ has not been applied strictly, systematically or universally. Some schools that are closer than 40 km to the mine have not been included in the Anglo American Platinum programmes, while others that are further away than 60km have been included. The support received by schools is fairly random and unsystematic, and probably reflects the relationships and levels of advocacy between Anglo American Platinum personnel, education authorities, and community leaders.

Although this allocation might have made sense and benefitted more schools in programmatic terms, from a research point of view the finding was disappointing: a ‘clean’ cut-off point design could not be applied; although with a bit more time and further work a ‘fuzzy’ discontinuity design (Cattaneo, Frandsen, & Titiunik, 2015; Lee & Lemieuxa, 2010) could be potentially explored for future replications of this impact evaluation.

b) The ‘proximity to mine effect’

As it was not possible to use cut-off point design in this circumstance, the next best option to calculate the net impact of the Anglo Platinum interventions was to utilise PSM in combination with DID method (see section 2.13). A counterfactual was thus artificially created using PSM technique. ‘NatEMIS’ was used as an identifier variable, and ‘50km from mine’ was used as the treatment dummy, in the matching operation, to estimate the intention-to-treat (ITT) effect (Duflo et al., 2008) of the mines on education outcomes in the surrounding schools.

Based on education theory (see chapter 3) and previous econometric analysis, the researchers decided to use the covariates listed in Table 14 for the matching process, as these were seen to have a greater determining effect on school results.

Table 14. *Covariates used for PSM*

EngAfr_HLang_learners	Quintile	Overall_Passes_2008
Per_black_learners	AverageFees	Bachelor_Passes_2008
Learner-Educator_ratio	SchoolType	Math_Enrol_2008
Science_Lab	Specialisation	Math_Passes_2008
Computer_Lab	Land_ownership	Physics_Enrol_2008
Media_Library	Learner-MST_Educator_ratio	Physics_Passes_2008
Multi-group_class	Dinaledi	English_Passes_2008
Total_FET_Learners	Any_Intervention_YES	English_Excel_2008
*Curriculum-variables[6]	*District-location-variables[7]	*Distance_Mine[8]

Matching control schools to treatment schools with regard to the general impact of the mine on learning outcomes was done using the nearest neighbour method through a logit regression, 0,2 calliper matching with a 1:2 ration without replacement. Table 15 illustrates the result of the PSM on '50 km from mine operations'.

Table 15. *Sample Sizes PSM Proximity to Mines*

	Control	Treated
All	873	655
Matched	382	293
Unmatched	491	362
Discarded	0	0

[6] These include: SgbApprovedCurriculum, CurriculumAvailable, TimeTable, CurriculumPlan, CurriculumTest, ImplementationSystems, AccesstoMaths, QualityofMaths, AccesstoScience, QualityofScience, IntegrateICT, EducatorAssistance.

[7] These include: Greaterdelareyville, Greatertaung, Kagisanomolopo, Kgetlengriver, Lebowakgomo, Lichtenburg, Madibeng, Mafikeng, Maquassihills, Matlosana, Mogalakwena, Mopani, Moretele, Moseskotaneeast, Capricorn, Potchefstroom, Polokwane, Sekhukhune, Rustenburg, Taledi, Vhembe, Waterberg, Zeerust.

[8] Used only in the next PSM on the impact of the Anglo programme.

Following the matching and the propensity score balancing, a treatment group of 293 schools and 382 control schools was created. To further reduce the difference between the groups based on unobserved variables, the DID method was introduced in the model.

A simple DID computation was made in SPSS, comparing the NSC results of each school in 2012 (post-test) with those of 2008 (pre-test). For example:

$$\boldsymbol{DID_MathPass} = (\boldsymbol{MathPass_{T2012}} - \boldsymbol{MathPass_{T2008}}) - (\boldsymbol{MathPass_{C2012}} - \boldsymbol{Math\,Pass_{C2008}}),$$

where

$\boldsymbol{MathPass_{T2012}}$ is the average percentage (%) of learners in the treatment school passing mathematics in NSC in 2012,

$\boldsymbol{MathPass_{C2008}}$ is the average percentage (%) of learners in the control school passing mathematics in NSC in 2008.

A DID value was computed for every school in both the treatment and control samples for all eight NSC outcomes of interest, in order to observe not only mean but also standard deviations of the results. Results were analysed against the following eight learning indicators in the NSC exams:

a) Mathematics enrolment

b) Mathematics passes (avove 30%)

c) Mathematics excel (above 50%)

d) Physics enrolment

e) Phystics passes (avove 30%)

f) Physics excel (above 50%)

g) Overall passes

h) Bachelor passes

An independent sample *t*-test was thus run to check if the mean difference between the programme and control sample was statistically significant. Table 16 shows the result.

Table 16. *Independent Samples Test on Groups Close to Mine and Far from the Mines*

		t-test for Equality of Means				
		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
DID_OveralPass	Equal variances assumed	-2.162	673	.031	-.04445	.02056
	Equal variances not assumed	-2.183	648.384	.029	-.04445	.02036
DID_Batchelor	Equal variances assumed	-2.038	673	.042	-.02100	.01030
	Equal variances not assumed	-2.077	663.122	.038	-.02100	.01011
DID_MathEnrol	Equal variances assumed	1.969	673	.049	.03928	.01995
	Equal variances not assumed	1.990	649.778	.047	.03928	.01974
DID_MathPass	Equal variances assumed	-1.859	673	.063	-.04684	.02519
	Equal variances not assumed	-1.884	654.988	.060	-.04684	.02486
DID_MathExcel	Equal variances assumed	-1.674	673	.095	-.02666	.01593
	Equal variances not assumed	-1.714	668.429	.087	-.02666	.01555
DID_PhysicsEnrol	Equal variances assumed	.557	673	.578	.00902	.01620
	Equal variances not assumed	.555	621.662	.579	.00902	.01625
DID_PhysicsPass	Equal variances assumed	-1.984	673	.048	-.05141	.02592
	Equal variances not assumed	-1.988	633.817	.047	-.05141	.02586
DID_PhysicsExcel	Equal variances assumed	-1.473	673	.141	-.02009	.01364
	Equal variances not assumed	-1.496	658.109	.135	-.02009	.01343

The results in Table 16 demonstrate that, based on a 90% confidence interval, most differences are statistically significant, therefore the proximity to the mine (within 50 km) did indeed have a small impact on the results.

In order to standardise the results, and allow for better comparison with other meta-analytical and educational literature, the raw results from the DID were converted into a 'Cohen's *d*' type effect size (Becker & Wu, 2007; Hombrados & Waddington, 2012; Lipsey & Wilson, 2001). To arrive with more

precision on the magnitude and direction of the impact, an effect size calculator was used by plugging in the mean, standard deviation, and sample sizes of both treatment and control, gathered through the *t*-test. The results are presented in Table 17.

Table 17. *Effect Size of 'Proximity to Mine Effect'*

	Treatment			Control			Effect Size
	mean	sd	n	mean	sd	n	
Maths Enrol	-0,0948	0,24548	293	-0,1341	0,26524	382	0,153232787
Maths Pass	0,0449	0,30572	293	0,0917	0,33804	382	-0,144474864
Maths Excel	-0,004	0,18344	293	0,0226	0,22026	382	-0,129886647
Physics Enrol	-0,0133	0,2111	293	-0,0224	0,20667	382	0,043688172
Physics Pass	-0,0141	0,33035	293	0,0373	0,33632	382	-0,154239456
Physics Excel	0,066	0,16393	293	0,0861	0,18405	382	-0,114632219
Overall Pass	0,0648	0,25388	293	0,1093	0,27274	382	-0,168350385
Overall Batch	0,0452	0,12167	293	0,0662	0,1405	382	-0,158535961

Both the *t*-test and the effect size calculation show that proximity of 50 km to a mine does indeed have a negative effect on overall learner passes, as well as mathematics and physics achievement. From Table 9 it also appears that the presence of the mine seems to have a small positive correlation with enrolments in mathematics and science exam subjects, but a negative correlation with pass rates. To translate these results in the education setting of Limpopo and North West, Table 18 shows how much schools close to the mines suffer percentage scores reduction in overall pass rates compared to other schools in the two provinces.

Table 18. *Handicap on Pass Rates Caused by Schools' Proximity to Mining Operations*

All schools in North West and Limpopo			Schools within 50 km from mines	
NSC Result 2012	Average	Standar Dev	Effect Size	Increase/Decrease
Overall Pass	60,5%	21%	-0,168350385	-3,5%
Bachelor Pass	14,5%	15%	-0,158535961	-2,5%
Maths Participation	52,5%	25%	0,153232787	4%
Physics Participation	39%	22%	0,043688172	1%
Maths Pass	45%	26%	-0,144474864	-4%
Physics Pass	57,5%	26%	-0,154239456	-4%
Maths Excel	18%	17%	-0,129886647	-2%
Physics Exel	12%	18%	-0,114632219	-2%

This further confirms the previous regression analysis (section 4.5.2) as well as the theory of the negative impact of mines on local schools that emerged during the qualitative fieldwork (section 4.5.1c).

c) Impact of Amplats education programme

The big question that remained was whether Anglo American Platinum's massive education programme implemented in the communities around its mining operations managed to reverse some of these trends and have a positive impact on the learning outcomes of the schools it supported. Did the Anglo American Platinum CED investments in the Limpopo and North West education sectors from 2009 to 2012 turn out to be effective? Once again, considering the circumstances, the best way to undertake this impact assessment was by combining PSM, DID, students' *t*-test, and Cohen's *d* effect size calculation, as undertaken above when measuring the impact of the proximity of the mine.

In the section on education interventions by Anglo American Platinum (section 4.3) it was shown that its education programme consisted of a number of different interventions (infrastructure projects, Radmaste teacher training, star schools, Radical Maths and Science Saturday and winter schools, etc.) implemented in different dosages in over 100 schools across Limpopo and North West. In order to maximise the potential effect of Anglo American Platinum's education programme, the researchers

decided to combine the various interventions into one mega Amplats programme dummy variable (indicated as Anglo_Yes), and use the entire population of 134 Anglo Platinum schools that received any type of treatment from the company to draw up the programme sample.

The treatment samples and control samples were created using the same covariates used above in the PSM for the '50 km from mine effect'. The main difference this time in the PSM was that the 'distance_mine' variable was added to the pool of covariates and the treatment became the dummy variable of 'Anglo_Yes'. The same PSM process as before was run by using the nearest neighbour method, logit regression, 0,2 caliper, but this time with a 1:3 ratio without replacement, and without discarding cases from the programme group as it was more difficult to find enough good matches for the Anglo Platinum schools. The result was the samples listed in Table 19.

Table 19. *Sample Sizes PSM on Overall Anglo Platinum Programme*

	Control	Treated
All	1393	134
Matched	159	90
Unmatched	822	44
Discarded	412	0

The same process outlined above (calculating the DID for all eight different NSC result outcomes, conducting an independent samples *t*-test for significance in the mean difference between programme and control group) was now done to assess the impact of the overall Anglo American Platinum education interventions package. Table 20 shows the results.

Table 20. *Independent Samples Test on Groups Receiving Support From Amplats and Those That Do Not*

		t-test for Equality of Means				
		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
DID_Math_Enrol	Equal variances assumed	.252	247	.801	.00840	.03335
	Equal variances not assumed	.244	166.953	.808	.00840	.03450
DID_Math_Pass	Equal variances assumed	-.067	247	.946	-.00262	.03896
	Equal variances not assumed	-.068	191.275	.946	-.00262	.03852
DID_Math_Excel	Equal variances assumed	-.493	247	.622	-.01233	.02500
	Equal variances not assumed	-.501	193.520	.617	-.01233	.02462
DID_Physics_Enrol	Equal variances assumed	-.891	247	.374	-.02187	.02453
	Equal variances not assumed	-.890	184.301	.374	-.02187	.02456
DID_Physics_Pass	Equal variances assumed	.092	247	.927	.00394	.04275
	Equal variances not assumed	.090	169.363	.929	.00394	.04401
DID_Physics_Excel	Equal variances assumed	-1.312	247	.191	-.02746	.02093
	Equal variances not assumed	-1.374	211.049	.171	-.02746	.01998
DID_Overal_Pass	Equal variances assumed	.453	247	.651	.01503	.03316
	Equal variances not assumed	.432	159.961	.666	.01503	.03479
DID_Overal_Batch	Equal variances assumed	.246	247	.806	.00383	.01560
	Equal variances not assumed	.251	197.708	.802	.00383	.01525

From the *t*-test it appears that none of the differences between the mean of the control group and programme group are statistically significant. To confirm this, Cohen's *d* for the eight NSC outcomes was calculated (see Table 21).

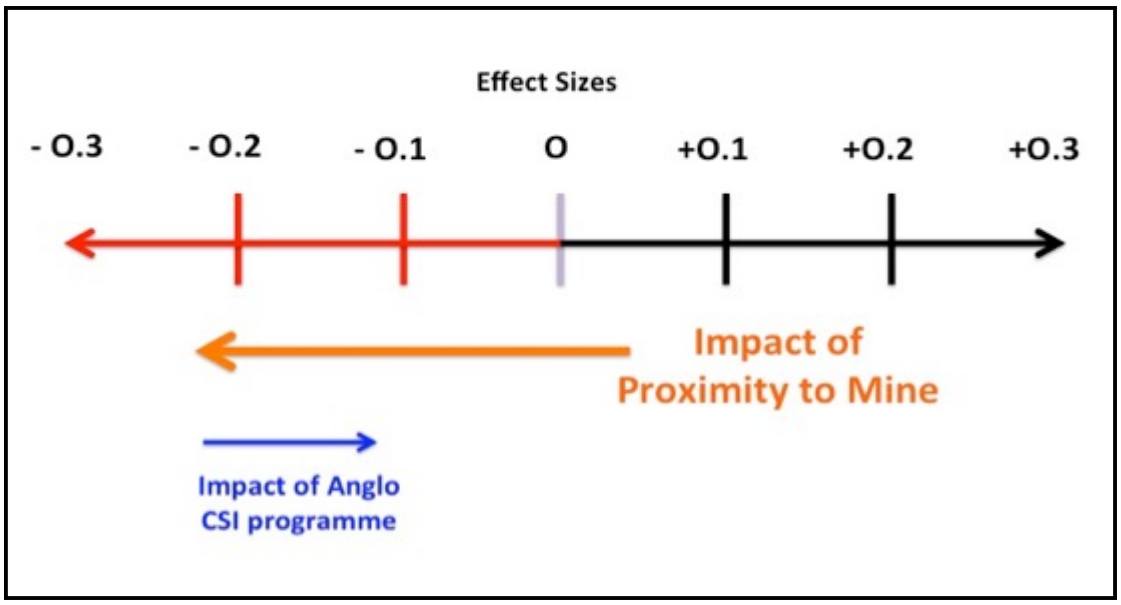
Table 21. *Effect Size of Overall Anglo Platinum Programme*

	Treatment			Control			Cohen's d
	mean	sd	n	mean	sd	N	
Math Enrol	-0,0893	0,27241	90	-0,0977	0,24113	159	0,033355863
Math Pass	0,0856	0,2876	90	0,0883	0,29961	159	-0,009178981
Math Excel	0,0101	0,18281	90	0,0225	0,19318	159	-0,065696675
Physics Enrol	-0,0099	0,18645	90	0,012	0,18573	159	-0,118224175
Physics Pass	0,0508	0,34564	90	0,0469	0,31127	159	0,012082886
Physics Excel	0,0698	0,14168	90	0,0973	0,16745	159	-0,174040523
Overall Pass	0,1229	0,27894	90	0,1079	0,23438	159	0,059919317
Overall Batch	0,0631	0,11209	90	0,0593	0,12163	159	0,032256631

The SMD effect size calculator also shows that the impact of Anglo American Platinum's education interventions on learning outcomes in treatment schools is not statistically significant (Cohen, 1988; Lipsey, 2007), and produces only a very tiny effect. In the few cases where the results are slightly more significant, the impact coefficient remains negative.

This does not necessarily mean that the Anglo American education intervention was *per se* ineffective. Rather in the light of the previous results on the impact of the schools' proximity to the mines, it highlights that the small effect of these programmes were not strong enough to counter-balance the large negative effect caused by the school's proximity to a mining operation. This concept can be graphically illustrated in Figure 37.

Figure 37. Positive and negative forces affecting learning outcomes.



Source: Authors own compilation.

As the results did not appear very positive, the analyst decided to scrutinise the Anglo American Platinum interventions to try to measure separately the individual components of the bigger education programme run by Anglo Platinum. One of the largest and more recent activities, with more data available, was the Radical Maths & Science Winter Camp for grade 12 learners, implemented in 2012 for 958 learners across the Amplats operation areas, which cost the company over ZAR 3.3 million (Amplats CED reports, 2013). The theory behind the intervention was that providing a high-quality two-week academic programme over the school holidays was going to increase the performance of the 12 grade learners in the NSC exams that they were going to write 4-5 months later. According to the service provider's report (Radical Maths & Science, 2012), learning assessments were conducted on all the learners participating

in the workshop before and after the training to observe how effective the programme had been in improving the test scores of participants in mathematics and physical sciences, among several other subjects[9].

Table 22. *Average Results From Pre and Post Test of Learners Attending 2012 Winter Enrichment Programme*

Area	Number of learners	Average Days attended	Average Mathematics Test Results - %			Average Physics Test Results - %		
			Pre	Post	Improv.	Pre	Post	Improv.
RUSTENBURG	170	10	46	69	22	16	48	33
THABAZIMBI	198	10	38	52	13	11	41	31
POLOKWANE	164	10	35	56	21	8	39	31
MOKOPANE	210	10	35	56	21	8	39	31
TWICKENHAM	216	10	45	58	12	15	41	25

Source: Radical Maths & Science, 2012

Table 22 shows that the 2012 winter camp assisted learners to improve their mathematics and physics test scores by an average of between 12 and 33 percentage points. One can thus conclude that the programme was very successful. As discussed in section 2.9 and as seen in other South African education evaluations (Fleisch et al., 2014), there are some serious problems in using such before-and-after assessments, without a proper counterfactual, which limits the validity of a scientific impact evaluation. From the results in Table 20 one can deduce that the Radical M&S Winter School had a reasonable impact on the knowledge and abilities of the learners who attended the 10-day intensive workshop. But then, why did this not translate in improvement in the overall school NSC results of that same year?

If in the evaluation of the Anglo Platinum interventions what is being measured is NSC passes at the school level (rather than learner level), it thus becomes important to assess whether enough learners from the Anglo Platinum-supported schools participated in the winter camp to cause enough impact on the entire school system. The researcher estimated the percentage of learners in treated schools by dividing the number of participants from each school in the Amplats 2012 winter school by the number of learners who enrolled in the same year (2012) in the NSC exam, as well as their enrolment in the specific subjects of mathematics and physical sciences[10]. The results from the 37 participating schools were as follows (Table 23):

[9] Aside from Mathematics and Physical Sciences, Life Science and Geography were also taught and tested at the 2012 Winter Camp.

[10] Programme data was made available by Radical Maths and Science and NSC exam enrolment data was provided by Umalusi.

Table 23. *Percentage of Learners in Treated Schools Attending 2012 Winter Schools*

Province	Institution_Name	District	Circuit	% of Total Grd12	% of Maths Grd12	% of Physcs Grd12
NW	GAOPOTLAKE SENIOR SECONDARY SCH.	MOSES KOTANE EAST	KGABO KWENA	31%	87%	93%
NW	KGALATLOWE HIGH SCHOOL	MOSES KOTANE WEST	KWENA	65%	100%	100%
NW	KWENATLASE	MOSES KOTANE EAST	MOGALE	35%	100%	100%
LP	MABOGOPEDI HIGH	WATERBERG	THABAZIMBI	43%	100%	80%
LP	LANGALIBALELE SECONDARY	WATERBERG	MAPELA	51%	100%	100%
LP	MAKGENENE HIGH	WATERBERG	MAPELA	3%	6%	6%
LP	MALEYA SENIOR SECONDARY	WATERBERG	MAPELA	11%	43%	33%
LP	MATANTA SECONDARY	WATERBERG	MAPELA	52%	82%	85%
LP	M.C. LANGA SECONDARY	WATERBERG	MAPELA	8%	29%	15%
LP	MMANTUTULE SECONDARY	WATERBERG	MAPELA	40%	51%	77%
LP	MMATEDU SECONDARY	WATERBERG	MAPELA	71%	94%	94%
LP	MPHUNYE SENIOR SECONDARY	WATERBERG	MAPELA	37%	84%	84%
LP	SEKOBA SECONDARY	WATERBERG	MAPELA	18%	40%	40%
LP	SERITARITA SECONDARY	WATERBERG	MAPELA	47%	54%	54%
LP	TJITJILA SECONDARY	WATERBERG	MAPELA	5%	22%	13%
LP	O.R. MABOTJA SECONDARY	CAPRICORN	SESHEGO	14%	100%	78%
LP	MAFOLOFOLO HIGH	CAPRICORN	DIMAMO	46%	65%	88%
LP	BOIKHUTSONG SENIOR SECONDARY	CAPRICORN	MOGODUMO	22%	50%	60%
LP	LESHIKISHIKI HIGH	CAPRICORN	MOGODUMO	3%	5%	5%
LP	MAHLOGEDI SECONDARY	CAPRICORN	MOGODUMO	71%	100%	100%
LP	MAHWIBITSWANE SECONDARY	CAPRICORN	MOGODUMO	24%	80%	95%
LP	RADIKGOMO SECONDARY	CAPRICORN	MOGODUMO	27%	75%	75%
LP	SEHLOLA HIGH	CAPRICORN	MOGODUMO	20%	75%	75%
LP	THOGOA SECONDARY	CAPRICORN	MOGODUMO	7%	17%	17%
LP	TSHEHLO SECONDARY	CAPRICORN	MOGODUMO	30%	43%	55%
LP	SHORWANE SECONDARY	GREATER SEKHUKHUNE	NGWAABE	31%	67%	67%
LP	GANGADZA SENIOR SECONDARY	GREATER SEKHUKHUNE	DILOKONG	100%	100%	100%
LP	LESAILANE SECONDARY	GREATER SEKHUKHUNE	DILOKONG	66%	92%	92%
LP	TEKANANG SECONDARY	GREATER SEKHUKHUNE	MOROKE	42%	100%	100%
LP	TSHIHLO SECONDARY	GREATER SEKHUKHUNE	MOROKE	43%	100%	100%
LP	MAKOPI SENIOR SECONDARY	GREATER SEKHUKHUNE	DRIEKOP	43%	100%	100%
LP	MAKGWALE SECONDARY	GREATER SEKHUKHUNE	NGWAABE	20%	91%	91%
LP	NGWANANGWATO SECONDARY	GREATER SEKHUKHUNE	NGWAABE	22%	33%	79%
LP	NKOTWANE SECONDARY	GREATER SEKHUKHUNE	NGWAABE	35%	100%	65%
LP	SENGANGE SENIOR SECONDARY	GREATER SEKHUKHUNE	NGWAABE	47%	94%	100%
LP	GOBETSE SECONDARY	GREATER SEKHUKHUNE	NGWAABE	23%	90%	100%
LP	ITIRELENG SECONDARY	WATERBERG	VAALWATER	100%	100%	100%
				Mean	**Mean**	**Mean**
				37%	**72%**	**73%**

On average, 37% of the grade 12 learners and about 72% of the learners who wrote the mathematics or physics NSC exams participated in the Radical Maths and Science Winter Camp in 2012, thus providing a big enough sample of participants to potentially have a significant effect on the overall school results at NSC exams in 2012, at least for the specific subject areas. In addition, 2012 should also be in theory the year in which the 2010 grade 10 learners involved in the Radical Maths and Science A-Team Saturday extra classes reached grade 12, thus adding additional effect to Anglo Platinum-treated schools.

As in previous calculations, PSM, DID, and *t*-test were run in SPSS using as treatment variable the dummy 'Wintercamp_2012', with the results shown in Tables 23 and 24.

Table 24. *Sample Sizes of PSM on 2012 Winter School*

	Control	Treated
All	1491	37
Matched	44	35
Unmatched	164	2
Discarded	1283	0

Table 25. *Independent Samples Test on Schools Attending 2012 Winter Schools Versus Those That Do Not*

		t-test for Equality of Means				
		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
DID_OveralPass	Equal variances assumed	.938	77	.351	.05957	.06349
	Equal variances not assumed	.914	63.902	.364	.05957	.06518
DID_Batchelor	Equal variances assumed	.350	77	.727	.00968	.02766
	Equal variances not assumed	.341	64.208	.734	.00968	.02837
DID_MathPass	Equal variances assumed	.818	77	.416	.05388	.06583
	Equal variances not assumed	.803	66.856	.425	.05388	.06706
DID_MathExcel	Equal variances assumed	.205	77	.838	.00841	.04094
	Equal variances not assumed	.205	72.260	.838	.00841	.04105
DID_PhysicsPass	Equal variances assumed	.808	77	.422	.05615	.06951
	Equal variances not assumed	.814	75.061	.418	.05615	.06894
DID_PhysicsExcel	Equal variances assumed	-1.232	77	.222	-.03381	.02744
	Equal variances not assumed	-1.237	74.071	.220	-.03381	.02733

Once again, when looking at the impact of the programme on all of the schools participating, the learner enrichment winter camp and the additional spillover from the A-team Saturday Classes did not produce a significant impact on overall and subject-specific pass rates in 2012, as would be expected. Another *t*-test

was subsequently run on the control and treated matched samples, this time removing four cases that represented schools with a low level of learners participating in the Anglo American Platinum programme (below 25%), and still no statistical significance emerged (Table 25).

To caution the reader, this does not necessarily indicate that the Radical Maths & Science Winter Camps were per se ineffective in raising learner pass-rates in beneficiary schools. The inconclusive results are simply due to the fact that sample sizes were too small to have statistical power to indicate the existence of a positive or negative impact of the programme on the learning results of the relatively few schools involved. One way to arrive at more significant results (because of larger sample sizes) and potentially also bigger impact would be to use learners instead of schools as the unit of analysis (see Lipsey et al., 2012). Although this exercise was attempted by the researcher, two major problems immediately arose. The first was an evaluation issue, where it would be difficult to establish an appropriate counterfactual, as the learners who attended the Radical M&S winter school were already the best performing ones, thus presenting a clear 'selection bias' towards the treated sample. The other problem was a data one, where Umalusi was not able to match in their databases the learner details provided from the limited information extracted from the Anglo American and service provider project reports. Furthermore, Umalusi was also hesitant to share learner-specific NSC results as there would be also confidentiality and ethical concerns in using such data without permission from the human subjects concerned, in this case the under-aged learners and their parents/guardians. Thus, the only data that could be used and the analysis that could be performed within this ex-post evaluation was only at the school level.

All the other interventions implemented between 2009 and 2012 by the company benefitted an even smaller number of participants from the various Anglo Platinum-treated schools (on average between 4% and 55%), and therefore they were even less likely to produce a significant impact on the whole school's results. In theory, some of the more systemic interventions should have a bigger impact on school-level outcomes. However, the infrastructure projects implemented in the period 2009–2012 were only a handful (seven science labs and three computer labs), often not functional (see Besharati, 2014), and not impactful enough to cause a direct correlation with learning outcomes (see previous section on the results of the econometric analysis).

The Radmaste programme, also implemented in 2012, could have been a promising intervention as it affected educators, who have a more systemic and long-term effect on school performance. The dosage was also reasonable as it affected 17-100% (61% on average) of the mathematics, science, and

technology (MST) teachers from 29 Anglo Platinum programme schools[11]. However, as discussed in earlier sections, due to financial and logistical reasons the programme could not be implemented according to the original plan and ended up being conducted with a smaller group and over a shorter timeframe in order to produce quick results for the 2012 NSC examinations (see more in Besharati, 2014). Considering that the Radmaste teacher development intervention occurred between July and September, that the academic year was near end and most of the teaching had already occurred, it had little time to cascade into improved teaching, learning, and results in the NSC exams of 2012. In future follow-ups to this study it would be useful to assess the impact of the Radmaste teacher-training programme against the 2013, 2014, and 2015 NSC results made available by Umalusi.

4.6 Conclusions

This chapter has illustrated the application of a great part of the evaluation framework presented in chapter 2 of this research, through an empirical case study that measured the impact of Anglo Platinum's education programme in North West and Limpopo provinces. The study utilised a mix of qualitative field-work with econometric, quasi-experimental design, and meta-analytical techniques to explore the effectiveness of the company's social investments around its mining locations. As part of the data collection, analysis, and discussion of the findings, a large spectrum of stakeholders from the mining and education sector were directly engaged in the research in a lively process of participatory applied research.

This concluding section will summarise the findings of the case study, the insights generated for the field of development evaluation, for South Africa's education sector, and the implications for Anglo Platinum future social investments. It will explain some of the policy and academic debates that were sparked by the results of the evaluation, discuss the elements of the evaluatoin framework which could not be applied, and thus explore avenues for potential follow-on to the above case study.

[11] The researcher divided the numbers of educators trained by Radmaste by the total number of Mathematics, Science and Technology (MST) teachers recorded by each school in that year through the Department of Education EMIS system, although data for some of the schools was missing.

4.6.1 Re-affirming the micro-macro evaluation paradox

Among other things, this case study has provided insights into some of the challenges and conundrums in development evaluation. Economists such as Mosley (1986), Boone (1996), and Picciotto (2006) have discussed in their evaluation studies the 'micro-macro paradox', elaborated previously in section 2.11 of this research. In the evaluation of the Anglo Platinum CED programmes instances of this paradox were also detected, where positive results from the assessment of individual participants at the micro-level do not necessarily translate into positive results at the macro-system and broader-sector level.

This is due to several challenges and limitations, particularly the lack of objectivity and scientific rigour often prevalent in evaluations of NGOs, corporates, and donors (Bamberger, 2009; Kruse, 2003; Masud & Yontcheva, 2005; Riddell & Kruse, 1997; Roche & Kelly, 2005). In this particular case, service providers such as Radmaste (2012) and Radical Maths & Science (2012) conducted the project-level impact assessments with learners and educators shortly after a workshop was conducted in 2011 and 2012. However, this does not necessarily mean that the knowledge gained will be retained and translated into results several months or years down the line, where many other forces are at play that might affect the final outcome (in this case, improved learning results). This is also consistent with the findings of Lipsey et al. (2012) that explain that lower effect sizes are expected in the education sector when assessing school-level interventions compared with the evaluation of small groups or one-on-one tutoring. NGOs, corporate investors, and development agencies often provide support to a small sample of individuals, who are not enough to make a significant change in the broader system.

A more serious weakness of project evaluations, such as the ones by Radmaste (2012) and Radical Maths & Science (2011, 2012) is that they were done lacking a counterfactual (or control) group. The improvement in the learning results of the participants is a normal trend and could have occurred regardless of the programme. Therefore, claims to causality and attribution made by a programme need to always be taken with great caution in the absence of an experimental setting (see also Fleisch et al., 2014). Also very common in development evaluation is the occurrence of the Hawthorne effect (Stand, 2000), where participants perform better simply because they are under observation, but these results do not endure in the long run once the evaluation and intervention are finished. Finally, one cannot exclude the possibility of a funding bias when service providers conduct evaluations of their own programmes for the donors who fund them, avoiding reporting on the negative aspects of their programmes and highlighting the positive results.

4.6.2 Bottom line for the social investor

Based on the records and documentation provided by the company, between 2009 and 2012 Anglo American Platinum implemented an education programme costing almost ZAR 100 million in support of the schools and communities around its mining operations in Limpopo and North West. The FET programme was implemented through a number of service providers (Radical Maths & Science, Radmaste Centre, Star Schools) and consisted of a mixture of learner enrichment programmes (through Saturday classes and winter camps), teacher training, and some infrastructure and facility upgrades. One of the prime development objectives of the programme was to increase pass rates and learning results, particularly in the critical subjects of mathematics and the sciences, for students living around the mines, to increase prospects of future engineers and technicians to be recruited from the local communities.

Although the few selected beneficiaries (learners and teachers) appreciated and benefited from the Amplats programmes, the interventions were too small to translate into significant gains for the general school systems of Limpopo and North West. Rigorous and scientific impact assessments of the Anglo American Platinum education investments did not reveal a statistically significant effect of the programme on the mathematics and science results of beneficiary schools. The small positive effect of Anglo Platinum's interventions was outweighed by the large negative effects that the mere presence of its mines had on the learning outcomes in schools surrounding the platinum operations. The study demonstrated that there is a strong negative correlation between the proximity of a school to a mine with a decline of mathematics and science pass rates. This phenomenon goes beyond the well-intended education interventions of the company and is rooted in larger psyco-social dynamics of communities living around mining areas (see more in section 4.5.1c).

The investments made between 2009 and 2012 therefore provided relatively small returns on development outcomes at the system level. Generally, the company tried to do too many things in too small doses, thus reducing its chances of making a deep and lasting impact on the education outcomes it tried to address. The most promising initiative was probably the educator training, but this was also done with too small a group and too little time to allow for a visible effect.

The evaluation highlighted that Anglo American's future community development initiatives and education investments need to be designed in a more effective and intelligent manner, aligned to government planning, based on evidence, empirical research, and lessons emerging from local experiences (see more in Besharati, 2014).

4.6.3 Participation and policy

The findings of the above case study provided further empirical evidence for many of the theories long-debated in South Africa education policy circles (Bloch, 2009; Fiske & Ladd, 2004; Fleisch, 2008; Motala et al., 2014; Taylor et al., 2013; Sayed et al., 2013; Spaull, 2013), and it revealed also some new insights with regard to the psycho-social effects that mines have on surrounding schools and rural communities. Following the release of the official report (Besharati, 2014), the study received significant media attention, with several television and radio interviews, and dozens of newspaper articles commenting on the findings of the research.

The intention of the evaluation approach proposed in chapter 2 was not to confine itself to a mere academic exercise, but to also provide relevant and real-time inputs into the policy and practice space (see section 2.16). The conceptual framework developed for this research follow an applied research orientation (Neuman, 2006) and included participatory techniques (Schafft & Greenwood, 2003) to engage project stakeholders. As such, even in the case study, subjects were partners in all aspects of the evaluative process, from data collection, to analysis, interpretation and application of findings (see more in Besharati, 2014).

As results started to emerge out of the research process, numerous stakeholder engagements were held with governmental, corporate, academic, non-governmental audiences to discuss the study findings and their implication for education, CSI, and mining policy. Between 2013 and 2014 a series of small (20 people) to large group (200 people) discussions were held with different stakeholders to present the results of the case study. Below are some of the most prominent of these interactions:

- Presentation for senior management of Umalusi Council for Quality Assurance in Higher Education (November, 2013)
- Presentation for Anglo American Platinum senior management (January, 2014)
- Academic Conversation at Wits School of Governance (January, 2014)
- Presentation for Maths and Science Learner Support Community (March, 2014)
- Presentation for evaluators and CSI practitioners at Tshikululu Social Investments (April, 2014)
- Mining and Education Stakeholder Policy Conference (May, 2014)

- Presentation at Department of Basic Education, Curriculum and Teacher Development Division Workshop (June, 2014)

During these various public presentations, stakeholders engaged in the sector had a chance to interrogate, validate, and criticise the findings of the research, but most important start policy dialogue with regard to future programming, investments, and reform required to address jointly the challenges of South Africa's education arena. Many of these topics have been previously discussed in chapter 3 and will be taken up again in the meta-analysis in chapter 5. Some of the issues highlighted in the policy and stakeholder debates listed above included:

- Improved coordination between public and private actors to address development challenges in South Africa and the role of the various partners.
- A reflection on the interventions which prove to work and those who have not worked over the past decade.
- The need to improve design of future programmes to maximise impact and results for targeted beneficiaries.
- An increased focus on teacher-centred interventions as deemed more sustainable, systemic, and long-term in impact.
- Renewed attention to be given to younger learners and lower grades of the schooling system for substantive change to occur.
- The value of learning and teaching support material in disadvantaged schools to enhance learning process.
- Understanding of the broader psycho-social dimensions of community development programming.
- The need for more evidence-based research and evidence-based planning of CSI projects and education policies, including the strengthening of monitoring and evaluation systems systems.

Among other things, the above case study has illustrated how the conceptual and methodological approach proposed in this research, can also be a useful and effective framework for stakeholder engagement and cross-fertilisation of academic research with public policy and development practice.

4.6.4 Follow-up work on the evaluation framework

A number of interesting subjects transpired as a result of this case study that would require more in depth exploration in subsequent evaluations and academic research. A full list of potential follow-up topics of research work stemming from this case study is available in the official SAIIA report (Besharati, 2014).

For what this research is concerned, however, the above case study has presented an effective application and adaptation of the evaluation framework proposed in chapter 2, to development planning and policy setting with an in-depth analysis of the impact of a public-private partnership to address education outcomes in the middle-income country context of South Africa. The case study has tackled the challenging issue of causality and attribution present in the evaluation space, and provided an independent, scientific, and empirical instrument to produce evidence on social phenomena and impact of development interventions. The framework has proven to be a useful instrument not only for academics but also for a range of funders and practitioners operating in South Africa's schooling sector.

The evaluation framework was successfully applied to estimate with fair precision the impact (or rather lack of impact) of the Anglo Platinum education investments to address learning outcomes in rural provinces in South Africa. It also generated new knowledge on unintended negative effects of mines on surrounding communities. In the process it made effective use of quasi-experimental approaches like DID, PSM and RDD, and piloted innovative geo-spatial approaches to impact evaluation utilising the inherent features of mining operations.

The methodological framework discussed in chapter 2, however, also includes a comparative dimension, which is implemented through tools such as CEA and multiple treatment meta-analysis. The results emerging from the Amplats education programme impact assessment were unfortunately not as encouraging and therefore not as useful for policy-makers and investors to look at in a comparative study with other education interventions implemented in similar context. This case study therefore did not utilise the comparative analysis methods which are part of the methodological package of this research.

The subsequent level of application of the evaluation framework will thus be illustrated better in the next chapter which will present a systematic review of all major evaluations and academic studies which utilised rigorous experimental and quasi-experimental methods to evaluate programmes or policies implemented in the last two decades to improve learning outcomes in South African schools. The above Amplats case study will thus be utilised as one of the primary impact evaluations included in the systematic review of South African education interventions. The use of meta-analysis and CEA to compare the effectiveness of interventions to improve learner achievement in South African public schools will be the illustrated in the next chapter of this research.

Chapter 5: Comparative meta-analysis of interventions to improve learners achievement in South African schools

5. 1. Introduction

The previous chapter illustrated the application of the conceptual and methodological framework to the evaluation of the effectiveness of the education programme of the corporate social investor, Anglo American, implemented in two Northern provinces of South Africa. Unfortunately, the full range of methods from the evaluation framework discussed in chapter 2, could not be tested. The previous case study, in fact, limited itself to the evaluation of the impact of a specific intervention but a subsequent comparative analysis was not possible, therefore the suggested methods for comparison such as multi-treatment meta-analysis and cost-effectiveness could not be applied. This was due to two main reasons, namely a) the results of the primary impact assessment were fairly small and disappointing; b) there were no other simultaneous impact evaluations of other similar interventions that could be compared against one another.

This chapter now tries to take the application of the evaluation framework a step further by piloting the approaches and methods discussed in section 2.18 and 2.19, for the comparison of the effectiveness of different types of interventions aimed at achieving the same development outcome. Remaining in the development policy context of South Africa's schooling sector, amply discussed in chapter 3, the following study takes a look at the numerous programmes and initiatives that have been implemented by different public and private institutions to address the challenges of quality education in South Africa (Bloch, 2009; Fleisch, 2008; Spaull, 2013). Acknowledging the limited evidence available in South Africa's school development literature on what works and what does not (Bloch, 2009; Motala and Pampallis, 2005; Sayed et al., 2013; Taylor et al., 2003), the following chapter conducts a systematic review of 127 effects of 31 education interventions implemented throughout the country in the past two decades – thus contributing new knowledge to the field of South Africa's education system.

The chapter appraises dozens of impact evaluations in terms of methodological quality and conducts an MTMA of rigorous experimental and quasi-experimental studies implemented in South

Africa's schooling sector. The comparative meta-analysis reveals what programmes and policies had the most impact on language, mathematics, and science learner results in public schools. It also explores the various contextual factors and internal design features that influence the magnitude of the effects reported in education impact studies. It also attempts to integrate a CEA, but this is met with numerous challenges, as will be seen later in section 5.8.

This chapter will utilise the methods, instruments, and approaches explained in chapter 2, for the comparative analysis of the effectiveness of different intervention options for South Africa's education policy-makers and investors. Section 5.2 will begin by summarising the South African education challenges, discussed earlier also in chapter 3, and provide an overview of the existing evidence and reviews already conducted in South Africa's education sector. Sections 5.3, 5.4, and 5.5 will clarify the objectives of the study, the conceptual and methodological framework, the criteria and parameters used in the systematic review, meta-analysis and CEA. Section 5.6 will provide the qualitative results of the systematic review of the various education programmes and their respective evaluations. Section 5.7 will present the findings of the statistical meta-analysis and section 5.8 will discuss the attempt made at cost-effectiveness comparison of the various interventions. The chapter will conclude by summarising the findings of the meta-analytical study and the results of the application of the comparative framework to South Africa's education policy, providing donors, government and managers with a synthesis of 20 years of learning and empirical evidence to inform new programming for the improvement of education outcomes in South Africa.

5.2. Background

5.2.1 Challenges of South Africa's education sector

Though a middle-income emerging economy, South Africa is a country that suffers from an extremely weak public school system. This exacerbates other national development challenges such as unemployment, poverty and inequality and racial divides. National learner assessments as well as international assessments such as SAQMEC (2007), PIRLS (2006), TIMSS (2003), have highlighted South Africa's consistent underperformance not only in comparison to the rest of the world but also to lower income countries on the African continent (Bloch, 2009; Spaull, 2013).

As elaborated in chapter 3, the historical legacies from the apartheid era created a racially segregated schooling system designed to limit the aspirations and potentials of the black majority. With the ushering in of the democratic dispensation in 1994, the government had the arduous task of undoing

decades of structural inequality and reconciling the previously 'white' and previously 'black' schools under the same schooling system (Fiske & Ladd, 2004). Numerous reforms were introduced such as the South African Schools Act (1996), the General and Further Education and Training Quality Assurance Act (2001). The national curriculum was revamped several times (1997, 2002, 2008), to follow an OBE model (Motala et al., 2014). The official South African school curriculum is aligned to the requirements of NSC exam, which is administered to all the learners in the country at the end of grade 12, by the independent education quality assurance agency of Umalusi.

The results from the NSC exam provide an important signal about learner knowledge and ability to enter further education and the work force (Kanjee, 2007). To cater for lower and mid-level learning assessments, the Department of Education administered in the past Systematic Evaluations (SEs), and in more recent years the Annual National Assessments (ANAs), which are standardised learner tests used in primary schools. Some provinces such as the Western Cape have also been implementing for many years, their own systematic learner assessment in the schools under their jurisdiction. All these provincial and national tests have been important tools for education improvement, 'internal accountability' (Elmore, 2008) and to monitor the quality of the overall schooling system of South Africa. Academics and research agencies such as the Joint Education Trust (JET), Human Sciences Research Council (HSRC), and Eric Schollar & Associates (ESA) have also developed specialised test instruments used in many education evaluations.

Notwithstanding twenty years of education reforms, the international as well as national assessments reveal that the majority of learners in previously black schools perform significantly lower than learners in previously white schools (Bloch, 2009; Spaull, 2013). The learning gaps are particularly acute in the subject areas of mathematics and sciences, which ironically are the gateway subjects needed to raise the engineers, technicians, accountants and financial analysts that the country desperately needs to sustain the main sectors of the economy (Besharati, 2014).

Numerous studies (Besharati, 2014; Botes & Mji, 2010; Taylor & Prinsloo, 2005; Simkins, 2010) have highlighted how language has been a serious obstacle to the learning of more complex mathematical and scientific concepts. The issue becomes more complicated when 83% of South African learners speak at home one of the nine vernacular African languages while switching to English or Afrikaans as their medium of instruction when they go to school. Linked to this is the findings (Van der Berg, 2008; Gustafsson, 2010; NPC, 2011; Taylor, 2011) that social-economic status (SES), such as household poverty, race, location, parent's educational background and other community factors, are also very highly correlated to learner performance.

Aside from the broader social-economic factors, systemic issues still remain and contribute to the weak education system of South Africa. Many schools, especially in the rural and poor parts of the

country still lack adequate infrastructure, facilities, materials, and resources to facilitate appropriate teaching and learning. Nonetheless, the most worrisome factor contributing to the poor learning levels has emerged to be the challenge with regard to the quantity and quality of the teachers (Christie et al., 2007; Deacon & Simkins, 2011; Taylor, 2011). The majority of South African educators were in fact trained under the post-1994 dispensation (CDE, 2011; DBE, 2013) and still suffer from very poor content knowledge as well as weak pedagogical skills (Spaull, 2011; Taylor, 2009). Absenteeism, accountability, productivity, and professionalism are serious concerns in South African schooling system (Chisholm et al., 2005; Langa & Du Toit, 2007; NPC, 2011; Taylor, 2011) that extend from teachers and school managers all the way to the educational authorities at district and provincial level.

Bloch (2009) has summarised the range of factors that contribute to persistent underperformance in learning outcomes in South Africa's public school system:

"…social disadvantage is reproduced across generations, where parents are often uneducated, relatively powerless and lack information… A range of issues affect teachers, from poor subject knowledge and teaching practices, to insufficient numbers in training and little performance evaluation… dysfunctional schools, accepting that schools mostly do not achieve acceptable outcomes, reinforced by confusion over OBE. Schools are badly managed and supported… Despite massive improvements, there are still huge backlogs: lack of libraries, labs and computers, not to mention poverty effects, from nutrition and AIDS orphans to gang violence… far stronger national intervention is needed to overcome inefficiencies as policy drops down to provincial delivery levels. District support systems and management in particular need to be fixed to give impetus to school-level improvement." (Bloch, 2009: 151)

5.2.2 The interventions: policies, programmes and investments

Notwithstanding the grim picture illustrated above, education has always been a very high priority in South Africa's national development planning processes (NPC, 2012), as this is closely linked to other social and economic problems affecting the country. South Africa contributes 5 -6% of its GDP every year to the education sector (OECD, 2008b). The majority of financing has been undertaken by the government, which utilises around 20% of national budget and 30-50% of provincial budgets every year on education spending (National Treasury, 2014). Among many reforms and enabling policies introduced since 1994, a no-fees regime applies to 70% of learners in the country and many benefit from a government-sponsored school feeding programme, as well as other schemes to promote universal access even to the poorest communities (see more in chapter 3). Education remains a high prerogative in

the Government's 2030 National Development Plan as well as the Department of Basic Education Strategic Action Plans

The private sector has also played a very prominent role in South Africa's education sector. Acknowledging the importance of a well-educated and capable young workforce for the success of businesses, SA's domestic corporate sector invests the biggest portion of its social investments (around R2-4 billion a year) in education (NBI, 2012; Trialogue, 2013). Corporate social investments in South Africa education sector surpass ten-fold the aid provided by the traditional donors, such as DFID, GIZ, UN, and World Bank (Besharati, 2015). Numerous public-private partnerships, such as the Joint Education Trust, the Business Trust, the National Business Initiative, and the latest National Education Collaboration Trust, have also been established to coordinate stakeholders in addressing the education challenges of the country.

Over the past two decades national and provincial governments, foreign donors, private companies, foundations, NGOs and universities, have implemented and piloted a range of diverse initiatives to improve the situation of South Africa's education sector. Where resources have been limited, interventions have been typically more narrow and focused. Government and bigger donors on the other hand have often recognised the complexity and the multi-layered deficiencies in the education system, thus have often taken a more comprehensive whole school development (WSD) programmes. These however have been fairly expensive initiatives and viewed by some as "scattered gun shot" (Bloch, 2009).

There are different ways of sub-dividing education interventions and several international reviews (Hattie, 2009; McEwan, 2015; Glewwe et al., 2011; Krishnaratne et al., 2013) have provided different typologies. Mouton et al. (2013) in their recent review have categorised school development programmes based on their target group (educational levels, phases, domains), geographic locations and delivery modes (support, resources, training, etc.). There are many other types of education interventions aimed at improving access, safety, and well-being at schools. Some like school feeding might achieve a number of simultaneous objectives such as incentivising enrolment/retention, improving nutrition, and thus cognitive abilities of learners. The focus of the following meta-analysis, however, are the interventions specifically aimed at improving learning outcomes and student achievement.

The intervention typology utilised in this meta-analysis will roughly follow the categories discussed in section 3.8 which elaborates in more detail the theoretical framework of the different interventions and the experiences that have emerged out of South Africa's education literature. It will also build on the the overall theory of change for South Africa's education sector, illustrated previously in Figure 23 in section 3.9. The types of interventions discussed in this chapter can thus be broadly summarised in the following manner:

a) **Learner-targeted support**: These are typically implemented by the private sector and civil society groups for a small group of selected learners. They consist of supplementary enrichment classes or tutoring by specialised teachers, often done after school hours, on weekends or during the school holidays. Another category of learner-focused interventions can be scholarships, bursaries and support to disadvantaged-but-promising learners to be placed in high-end schools or special courses. Learner-targeted initiatives tend to be fairly expensive, and are criticised for not being sustainable and benefitting only a small group of individuals.

b) **Teacher-centered initiatives**: acknowledging the central role that educators play in the learning process, many interventions are implemented to address teacher quantity, quality and performance. This can include the provision of additional teachers and assistants for a short or long period of time. Teacher development is usually divided into pre-service (PRESET) training and in-service training (INSET). The later consists of upgrading of qualifications, knowledge or skill-sets through different models of training implemented over weekends/evenings, school holidays or several months of block-release college/university programmes. Effective teacher development programmes often also have a component of classroom visits, coaching, monitoring and experience sharing between participants.

c) **Learning and Teaching Study Material (LTSM)**: These are materials, technology and other resources provided to schools to enhance teaching and learning experience. Examples of such can be textbooks, workbooks, study guides, lesson plans, science-kits, computers, multi-media accessories, educational toys, reading books, etc.

d) **Management and Governance:** These are interventions aimed at strengthening school management and leadership, often done in the form of training and coaching of principals and school management teams (SMTs). These often are geared to putting new systems in place to improve school functionality, and promoting accountability, efficiency and good governance. These can include also the introduction of specific incentives and prizes for achievement and performance.

e) **Infrastructure and facilities:** These are the classical *hardware* investments such as building, upgrading, expanding and refurbishing schools, classrooms, administration blocks, toilets, science labs, sporting facilities and libraries. Although not always as obvious as some of the

above interventions, the atmosphere, cleanliness, size, lighting, order and other environmental aspects of a school can have an impact on the learning processes. Such school infrastructure projects are often commissioned to external companies or given to the school to manage through a donation or conditional grant.

f) **Structural reforms, policies and incentives:** These are interventions aimed at improving the broader education system that support the schools. These can include programmes for district development, introduction of specific language policy, systematisation of learner assessments, performance rewards, decentralisation of functions, and other structural changes to improve management and accountability. These are normally interventions done by or with the government's Department of Education at national and sub-national level.

g) **Community/family involvement**: These acknowledge the important role parents and community play in the learning process inside and outside the formal schooling process. Such initiatives extend programme components to other non-school stakeholders and broader community actors. These include also interventions aimed at strengthening school governing boards (SGBs), which often include parents and other local authorities.

h) **Integrated School Development**: These are complex multi-layered programmes that combine three or more of the above interventions into a holistic strategy to improve school functionality and ultimately learning results. These are usually organised in the form of whole school improvement models that are built on prior research and specific theoretical frameworks, piloted in few areas and then scaled to many schools with large injections of funding and cooperation of various partners.

As a caveat, the above eight categories of interventions have been defined for the purpose of conducting a meta-analytical review of South Africa's education sector. It needs to be acknowledged, however, that these typologies are inevitably limiting, subjective and reductionist as these divisions are not as clear-cut as they appear to be, as most education programmes and policies operate in a complex manner and are often the combination of different interventios and approaches, and the interplay between them. The classification of interventions used in studies like this, can greatly affect the final results and conclusions of sysematic reviews, as was found also by Evans and Popova (2015) 'review of education reviews', discussed in more detail later in the chapter.

5.2.3 Growth of evidence-based education in South Africa

"The single most important lesson learnt about schooling by researchers, non- governmental organisation (NGOs) and government, in a decade of activity in schooling, is that it is a social phenomenon of immense complexity, opaque to the best-intentioned interventions based on the most self-evidently righteous explanations... it would seem that no one knows quite why the best efforts have produced so little change, or quite why schooling outcomes at levels other than matric, despite our best efforts, seem to have declined even further." (Taylor, Muller and Vinjevold, 2003, p. 128)

Section 2.17 discussed the rise of systematic reviews as advanced research approaches to synthesise existing evidence on the effectiveness of different treatments and interventions. Systematic reviews, which have dominated the medical, psychology and health sciences since the 1970s have been spearheaded by organisations such as the Cochrane Collaboration. Through the work of the Campbell Collaboration systematic reviews have been expanded also to the social sciences such as criminology, social welfare, and education. More recently they have been extended also to the development field by London-based institutions such as 3ie, ODI, and DFID. Systematic reviews have been used by the British government since the 1990s as a tool for evidence-based policy-making (Davies, 2012). The purpose of systematic reviews is to summarise existing knowledge of what works and how effective treatments are, and explore how interventions work on different populations and contexts (Dehejia, 2013). Randomised experiments as well as systematic reviews are growing rapidly in the field of education as part of a movement towards more evidence-based research and policy.

Similarly, in South Africa, government agencies, private philanthropies and academia have tried to address the deficiencies in the education sector by reviewing the evidence emerging from the numerous policies and programmes implemented over the years. The Joint Education Trust, a non-for-profit education agency established in the early 1990s, has been one of the forerunners in the use of research and evaluation to inform education programming in South Africa (Fleisch, 2015).

Zenex Foundation, a prominent education grant-maker committed to evidence-based programming, commissioned prolific evaluators such as Roberts and Schollar (2006) to conduct a meta-evaluation on the interventions aimed at improving language, mathematics and science learning outcomes in South Africa. A similar follow-up review on the Zenex-funded school programmes was conducted again by the same authors in 2011. Building on some of their previous scholarly work (Fleisch, 2008; Taylor et al., 2003), Taylor, Fleisch, and Schindler (2008) conducted a review for the Presidency on the changes in South Africa's education system since 1994. In the same year the Organisation for Economic

Cooperation and Development (OECD) conducted a large review of national education policies implemented in South Africa (OECD, 2008).

In more recent years, Sayed et al. (2013) published a review of some of most prominent programmes to improve the quality of education in South Africa, compiling lessons learnt from the Education Quality Improvement Partnership (EQUIP), the District Development Support Programme (DDSP), IMBEWU I and II, Quality Learning Programme (QLP), the Learning for Living (LfL), the Integrated Education Programme (IEP) and the Khanyisa School Development Programme. Many of these initiatives will be discussed in more detail also in this chapter.

One of the most comprehensive review efforts in South Africa's education literature was conducted by Mouton et al. (2013) in order to inform the Zenex Foundation of the next phase of school development programming. The review synthesised evidence from hundreds of evaluation reports, journal articles, books, and previous meta-evaluations that looked at the effectiveness of education development programmes. The report provided the client (Zenex) a framework to analyse the literature and make evidence-based decisions on their future school development programme. As part of the review Mouton et al. (2013) provided an evaluation framework which illustrated the range of studies that spanned from experimental, theory-based evaluations (TBE) to case studies, which he particularly favoured.

Although containing many elements of systematic reviews, Mouton et al.'s review (2013) did not follow strictly the systematic review protocols outlined by the Campbell Collaboration or by the PRISMA (www.prisma-statement.org) for systematic reviews. Mouton's review conducted a certain level of quality appraisal of the primary studies focusing on the ones marked as 'credible' and 'moderately credible' to draw out evidence and lessons to inform future programming. The approach was mainly of a 'narrative review' (see section 2.17), which provided qualitative analysis, depth, texture, context and elaboration on the 'mechanisms of change' of school development in South Africa. The reviewers however did not provide an empirical comparator between impact results (i.e., effect sizes) from the different evaluation studies, therefore making it more difficult for policy-makers and investors to assess the best among the competing options and many good interventions to choose from. In order to take Mouton et al. (2013) review a step further, the following chapter coducts a systematic meta-analysis of the major education interventions implemented in South Africa, integrating statistical analysis of effect sizes with qualitative analysis of the interventions and the evaluations conducted on them.

5.3. Objective

The following meta-analysis builds on previous education reviews, by offering a standardised and comparative framework to assess results of impact evaluations of interventions to improve learner achievement in South African public schools. It will synthesise the experimental literature produced post-1994 and highlight some of the programmes and policies that have been most effective in improving language, mathematics and other learning outcomes. It will explore the internal features, external factors and specific context in which such interventions have yielded the most impact. The systematic review provides useful information for education policy by highlighting successful programmes that can be capitalised, replicated, and scaled up by both government and private investors to achieve improved learning outcomes in South Africa's schooling system.

5.4. Conceptual Framework

5.4.1 Education evaluations in South Africa

The following section returns to some of the problems and constraints of development evaluation, discussed in section 2.8, but now with specific focus on the evaluations and studies conducted in South Africa's education sector. In the past two decades there has been an abundance of projects and initiatives implemented by different institutions to improve learning results in South African schools. Each of these interventions may have been evaluated with different methods and approaches, with different degrees of rigour, and thus successes or the failures of each programme have been reported in a very heterogeneous manner.

A number of evaluations undertaken in South Africa's education sector have utilised the case study approach (Mouton et al., 2013), often due to the limitations with data, methodological expertise, budget, time and context in which they were implemented. Although in many cases such studies were the only option, these still carried value and contributed knowledge to the field. Nonetheless, the bulk of South Africa's education evaluations rely heavily on qualitative approaches, which on one hand provide depth, context and texture, but on the other hand by nature are more subjective and susceptible to bias and personal experiences of the evaluator and respondents (Bamberger, 2009). This poses questions with regard to the independence and scientific validity of many of the agency commissioned evaluations (Riddell & Kruse, 1997; Roche and Kelly, 2005). Evaluations capturing no or negative results of

interventions often go unpublished or are concealed from public debate; while positive evaluations are used by donor and implementing agencies as 'marketing material' (Deaton, 2010). Because of their vast heterogeneity, qualitative studies are also more difficult to compare between each other and assess against a common scale.

The quantitative education evaluations have utilised statistical indicators to empirically measure and compare change in outcomes. Using a results-based language (Kusek & Khatouri, 2006), the tendency of many education evaluations, especially the ones commissioned by corporate social investors (Trialogue, 2010), focus on the reporting of inputs (how much was spent), activities (what was done), and occasionally outputs of a programme (i.e. number of beneficiaries). Few evaluations are done looking at the impact of initiatives at the level of outcomes, such as real changes (behavioural, institutional, social-economic) in the recipient population.

When it comes to education interventions the development outcome most frequently used to measure progress is 'learner achievement' (Schollar, 2015). This is traditionally measured through the indicators of 'test scores' or 'pass rates'. In South Africa, most learning assessments are designed and implemented by experienced evaluation and research organisations such as the HSRC, JET and ESA. Some researchers however prefer to use state-run standardised universal learner assessments, such as NSC and ANA results, discussed earlier.

Like in the rest of the developing world (Fuller & Clarke, 1994; Hanushek, 1995), most education evaluations in South Africa are non-experimental. Many studies popularly referred to as impact evaluations are actually case studies where beneficiary outcomes are observed only after the intervention. More advanced evaluations conducted baselines and are able to measure change before and after intervention, but these are relatively few (Bamberger et al., 2010). Some studies have gone to the extent of creating a comparison group, and in best-case scenarios have measured pre- and post- results of both programme and control group. A range of literature exists also in South Africa (i.e., Case & Deaton, 1999; Fleisch, 2006; Gustafsson, 2007; Hunt, Schoer, Nthuli, Rankin, & Sebastiao, 2010; Simkins & Paterson, 2005; Taylor & Yu, 2009; van der Berg, 2008) that utilise more complex econometric models to control for different variables and estimate effects of interventions, school and community factors on learning outcomes.

There are nonetheless still some complications with the above non-experimental approaches. Intuitively when measuring the impact of an education intervention one compares test scores of beneficiary populations before and after the intervention, and if there is a significant improvement one concludes that the intervention was successful and had an impact. In real life however there are a lot of other internal and external factors (social, economic, institutional, environmental) as well as other interventions of different parties that may affect the learning outcomes of a particular group of people

(Glewwe & Kremer, 2006; Jacoby, 2002). The great challenge of impact evaluation in social science and development policy is measuring attribution (Mouton et al., 2013; White, 2005) and causality of the intervention (Glewwe et al., 2004; McEwan, 2014). How do we isolate the effects of one specific intervention from other forces, processes and events affecting the same population (Johnson & Lamdany, 2005; Rossi & Freeman, 2001)?

Credible impact evaluations address the issue of causality through the establishment of the 'counterfactual' (White, 2009), in other words comparing a hypothetical situation where the same beneficiaries are observed with and without the programme. In real life this is impossible, but in evaluation practice this is achieved by creating a plausible control group that is identical (or at least very similar) to the treatment group. Assuming that all the other endogenous and exogenous factors remain the same throughout the process, the raw difference between the programme group and the control group indicates the impact (or the net effect) that an intervention produced.

The most convincing way to establish a counterfactual is through random assignment (Taylor, 2015), as this will lead to two identical groups in both observable and unobservable characteristics (Baker, 2000). Thus RCTs are considered the 'gold standard' (Levy, 2009) for empirical and counterfactual evaluation. There has therefore been a rise in the use of experimental approaches for the evaluation of education interventions in developing countries (McEwan, 2014; Dufloet al., 2008) and also in South Africa (i.e., Fleisch et al., 2010; Louw et al., 2008; Schollar, 2015; Taylor & Watson, 2015)

As discussed in section 2.13, in the social sciences and hence also in education, RCTs are very difficult to implement. They tend to be expensive and require a high level of expertise. While they are strong in establishing internal validity they are weak in external validity as the results are very context-specific (Deaton, 2010). Randomised control trials often face ethical, political, and practical constraints, as they need to be built in at design stage in the beginning of an intervention. Impact evaluations of South African government policies (such as curriculum, institutional arrangements, language policy, etc.) have also been difficult to perform as interventions have often been rolled out in all schools of the country, therefore not leaving much space for a counterfactual. This challenge was encountered, for instance, by Coetzee & Van der Berg (2012) as they attempted to evaluate the impact of South Africa's school nutrition programme for the DBE. Smaller projects, implemented in specific geographic areas and with large but still limited number of beneficiaries, where similar control schools (or learners) with similar characteristics can be identified, provide better conditions for the implementation of experimental evaluations.

Because of the above constraints, what has dominated much of the impact evaluations in South Africa's education sector has been the use of quasi-experimental approaches - discussed at length in section 2.13. Quasi-experiments are much easier and more commonly seen in the education literature, and

thus have been the choice of many of South Africa's education evaluators. As will be illustrated in more detail later in this chapter, by and large the most common counterfactual evaluations used in South Africa's education sector have been the combination of DID designs (Jacoby, 2002; Romero & Noble, 2008) with some form of matching techniques - including PSM (Besharati, 2014; Hobden & Hobden, 2009). Other South African researchers have tried to emulate RCTs, by using other evaluation approaches such as natural experiments (Gustafsson et al., 2013) or cut-off point design (Fleisch & Schoer, 2014).

As previously discussed in chapter 2 (Dehejia, 2013; Mouton et al., 2013; Shadish & Cook, 2009; Smith & Glass, 1977), if performed properly, such quasi-experimental evaluations can yield just as good impact estimates as randomised experiments. However, in the case of both RCTs and quasi-experiments, the assumptions and implementation need to be carefully checked so as to assess the plausibility of the counterfactual provided in the evaluation. Each study needs to be appraised in terms of methodological rigour, internal validity, and the presence of potential biases. Hence, the meta-analysis would contain a clear quality process for the appraisal of primary studies, which will be explained in more detail in later sections.

5.4.2 Standardising impact measures across studies

Experimental and quasi-experimental studies provide evidence on the impact of interventions. But most primary studies are concerned with whether there is a statistically significant effect, but less on how much this effect actually is. The studies that report the magnitude of the impact usually do so by providing an estimate measure in the form of ATE, average treatment on the treated (ATT) or intention to treat (ITT) (Duflo et al., 2008). But can we compare such impact measures between studies and interventions?

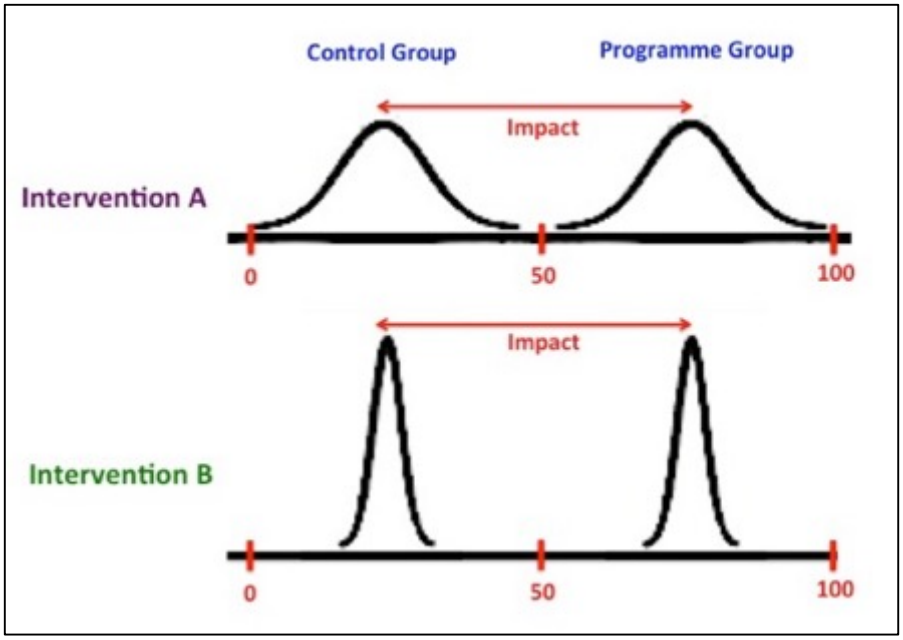
The problem arises when different evaluations utilised different scales, instruments, approaches, and systems of measurement. Some South African researchers may have used pass rates (above 30%) for 12 grade NSC exam as the indication of learner achievement (Blum et al., 2010; Besharati, 2014), while other evaluations might have used a self-made learner test which yields a maximum score of 50 points (Schollar, 2005). Some studies might have estimated the impact through the beta-coefficient of a multivariate regression (Gustafsson & Taylor, 2013) while others have estimated the impact through a simple comparison of means between programme and control groups (Schollar, 2002). These measurement tools are clearly different therefore would not be appropriate to compare the results of one study against another at face-value.

When conducting a systematic review of different evaluations, it is important to standardise the results into a common impact measure for fair comparisons. This statistic is referred to as *effect size*.

Effect sizes provide analysts and policy-makers with a common translation for the direction and magnitude of the impact of interventions (Hombrados & Waddington, 2012).

Some of the earlier education impact evaluations in South Africa would measure the effect size of an intervention by comparing the mean scores of the programme and control groups. But such measures would not be precise as they did not take into account the standard deviations of the two groups when analysing the mean difference. Figure 38 illustrates this point more vividly.

Figure 38. Graphical illustration of standardised mean difference.



Source: Authors own compilation (2014)

From Figure 38, one can see that impact assessments of two different interventions can report the same mean difference, however intervention B has clearly a more drastic effect on the population than intervention A. When comparing different types of impacts across studies with different instruments and scales it is thus important to measure the change in standard deviations between the two groups. Effect size thus indicates how many standard deviations separate the treatment and control group (TEA, 2014). For this reason, in much of the education economics literature, impact is often translated through the reporting of the *z* -scores.

There are several types of effect sizes (odds ratio, glass delta, correlation coefficient *r*, proportion overlap U_3) that can be transformed from one into another. One of the most commonly used effect sizes,

particularly when utilising continuous variables (such as learner test scores and pass rates) is Cohen's *d*, also known as SMD. This effect size is computed as shown in [14]:

$$\boldsymbol{SMD = \frac{X_t - X_c}{Std_{p\,(t+c)}}}, \textbf{ (14)}$$

where $\boldsymbol{X_t}$ is the mean score of the treated group, $\boldsymbol{X_c}$ is the mean score of the control group and $\boldsymbol{Std_{p\,(t+c)}}$ is the pooled standard deviation of the scores of all participants.

In our South African meta-analysis, most of the primary impact evaluations utilised the DID method, therefore statistics were provided for pre and post for both control and treated group. In such instances the formula expands to [15]:

$$\boldsymbol{SMD = \frac{(X_{post} - X_{pre})_t - (X_{post} - X_{pre})_c}{Std_{p\,(t+c)(post+pre)}}} \textbf{ (15)}$$

Whether large or small, effect sizes may still be or not be statistically significant. Effect sizes are still in fact just estimates therefore they are always accompanied by other statistics such as standard error, variances and confidence intervals (usually set at 95%). Sample size has a large influence in estimating the precision of the effect size. Therefore, the larger the sample the smaller the standard error will be and the confidence interval will have a shorter range. The way sampling is done, whether random or purposeful, as well as the different layers in which sampling is done (learner, class, and school level), makes a difference in the standard error (Bloom, Richburg-Hayes, & Black, 2007; McEwan, 2014; Taylor, 2015). Thus wherever clustered sampling occurs, effect sizes and standard errors will be adjusted accordingly.

Hedges (1981) devised a formula to correct Cohen's *d* estimates when sample sizes are small, therefore in this meta-analysis Hedge's *g* will also be used as the main effect size to compare the various studies and interventions. In order to calculate Cohen's *d* and Hedge's *g*, the reviewer needs to extract some basic statistics from the primary studies such as sample size (for the learners, classes and schools), and the means and standard deviations for the pre- and post-scores of both programme and control groups. The methods for calculating the different effect sizes and other accompanying statistics (i.e. confidence intervals) are illustrated in detail by authors such as Lipsey and Wilson (2001), Borenstein, Hedges, Higgins, and Rothstein (2009). There are also various software packages such as RevMan, Meta XL, Wilson's ES calculator, which assist with these calculations and conversions. The following study will utilise a statistical package extensively utilised in the meta-analytical field known as CMA.

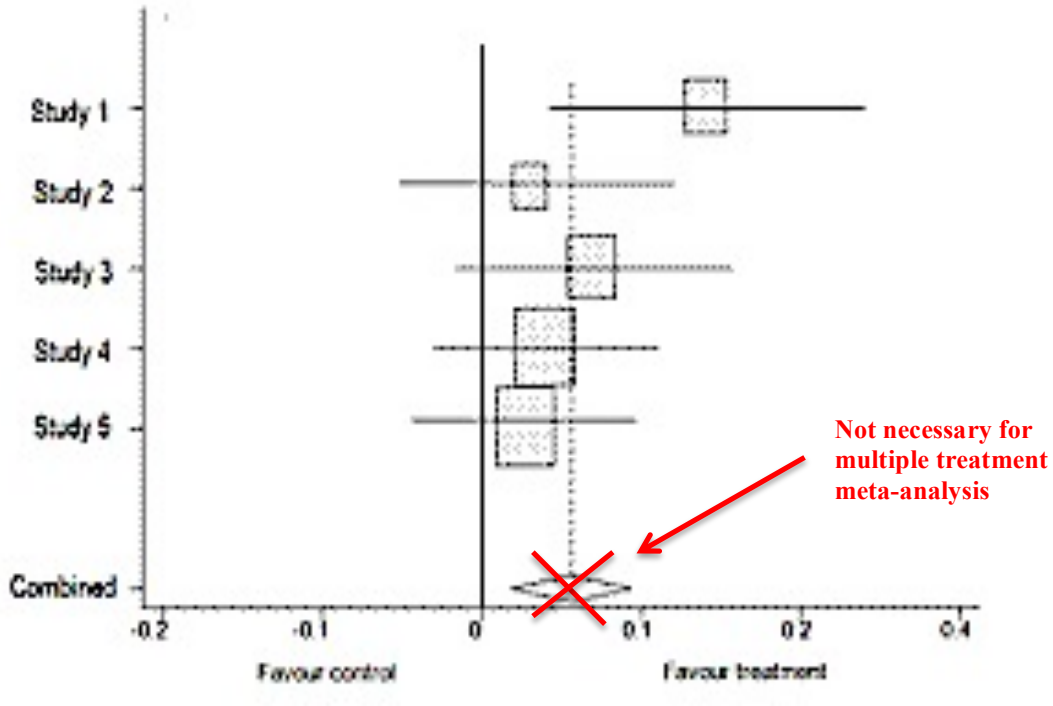
5.4.3 Multiple treatment meta-analysis (MTMA)

Section 2.18 has introduced the methods of meta-analysis as the quantitative arm of systematic reviews. Randomised control trials are poor in external validity however when evidence from various experiments is brought together in a meta-analysis, external validity also increases. Traditional meta-analyses tries to synthesise, observe patterns and relations between different studies, to arrive at higher forms of evidence. Through fixed and random effects models, results from different studies are combined together, weighted based on sample errors, in order to yield to a more precise effect size for a specific treatment or intervention.

For the current enquiry aimed at comparing effectiveness of different types of education programmes, specific types of meta-analytical approaches such as MTMA, MTC, and NMA are more appropriately used. All of these are extension of traditional meta-analysis elaborated by academics such as Caldwell (2005), Lu and Ades (2006), and Salanti (2007). As opposed to standard meta-analysis, multi-treatment meta-analysis assesses and compares different types of interventions aimed at achieving the same outcome (i.e., improved learner results).

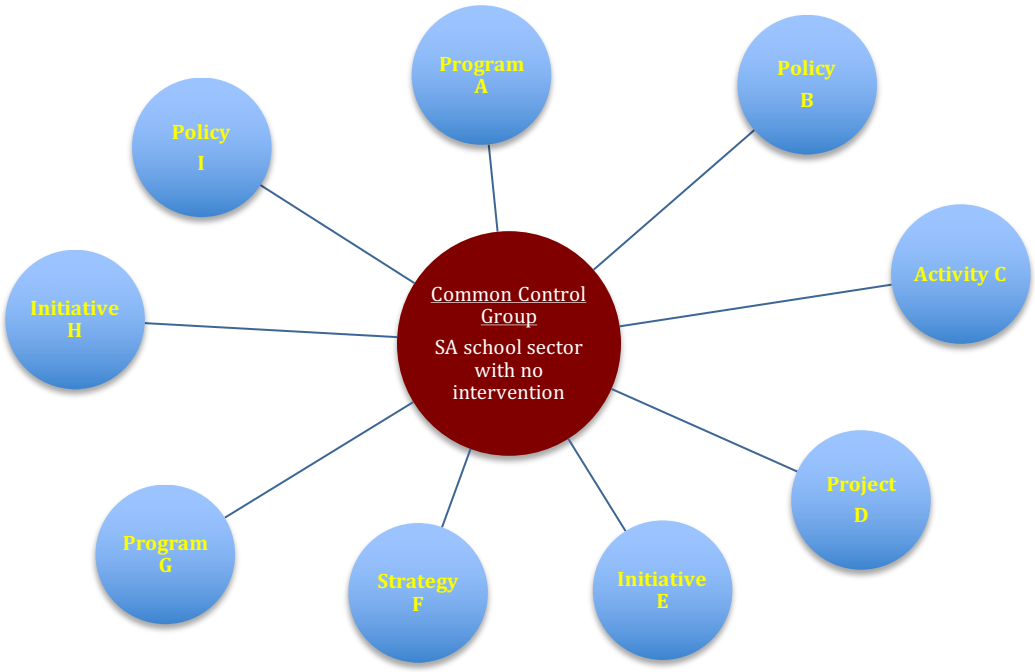
This method helps to answer policy questions and assists decision-makers when there are various options on what investments to make and which interventions to favour. In Multiple Treatment Meta-Analysis (which will be used in this study) heterogeneity is expected, therefore a random effects model, that excludes the calculation of the average pooled effect size, is what will be used. The results of a meta-analysis are traditionally illustrated through a forest plot (Figure 39) that provides decision-makers with an intuitive graphical illustration of different effect sizes and confidence interval of the various included studies. This allows for an easy assessment of the best intervention to address a particular outcome of interest.

Figure 39. Illustration of forest plot in a comparative meta-analysis.



In individual impact evaluations different interventions are tested against different populations, which can be more or less responsive to the treatment (Cranney et al., 2002), depending on context, time and other factors. One key feature of MTMA is that it tries to emulate a large multi-treatment experiment by utilising a common control group. Comparisons are made through observational means, by assessing one intervention in relationship to another intervention or a null (placebo) situation. This model is based on the assumption that the comparator used to assess the effectiveness of the various interventions is somewhat similar and consistent. In our particular case the common control group is South African public schools running business as usual and affected by the same social, economic, political and environmental factors present in the country. Below is a graphic illustration (Figure 40) of the multiple treatment meta-analysis that was conducted throughout this study.

Figure 40. Graphical illustration of multiple-treatment meta-analysis.



Source: Author's own compilation (2014)

Although the various experimental evaluations have occurred in the same country's educational system, it needs to be also acknowledged that the context could have varied based on the provincial/local characteristics and the historical period in which the intervention and the evaluation occurred.

Like in any other systematic review and meta-analyses, this South African study will proceed to undertake the following key steps as part of its analytical process:

1) choosing the criteria by which studies are included and excluded from the meta-analysis;
2) exhaustive search, review and screening of studies based on the selection criteria;
3) quality appraisal of the primary studies that match the original selection criteria;
4) calculation of effect sizes and confidence intervals for each study, intervention and outcome of interest; and
5) exploration of heterogeneity between studies and interventions, through moderator analysis.

These five steps will be elaborated in more detail in the succeeding sections. In line with systematic review standard practice (Campbell, 2015), the following South African education meta-analysis was

conducted by multiple reviewers: the primary researcher, Neissan Besharati, and two research assistants, Mpho Litha and Khotso Tsotsotso.

5.4.4 Integration of cost-effectiveness analysis

The following study endeavoured to also integrate cost-effectiveness measures in the meta-analysis, to provide a better comparative framework for policy-makers and investors on which intervention to favour for future programming. In their seminal book Levin and McEwan (2001) have elaborated on the different forms and approaches to CEA. Recently many researchers have started to integrate additional cost-effectiveness components to their impact evaluations (Duflo et al., 2012; Kremer et al., 2013; Taylor & Watson, 2015). The effort even in this meta-analysis, was to use CER as an additional parameter to be used for comparison between interventions. The CER for each intervention was calculated as follows (16):

$$\boldsymbol{CER} = \frac{(\boldsymbol{Total\ Cost})/\ (\#\ \boldsymbol{learners})}{\boldsymbol{Effect\ Size}} \textbf{(16)}$$

This analysis would thus provide a South African rand value (ZAR), adjusted for inflation and yearly discount, of how much the intervention would cost for every standard deviation change (effect size), adjusted for the number of learners benefitting from the programme. The number of years the programme would be running would influence both the costs and the total participants recorded in the programme.

The quantitative impact value of the cost-effectiveness equation would be the effect size of the intervention, calculated through the meta-analytical techniques previously discussed such as SMD and Hedge's *g*. The information about the number of beneficiaries benefitting from the treatment as well as the overall costs was extracted from the official programme reports and evaluations reviewed.

Where the information was not present in the documentation, the reviewer contacted the authors of the original evaluations, and, if necessary, the managers and donors of the interventions, who are all mostly still based in South Africa. The 'ingredients methods' (Chambers & Parrish, 1994; Dhaliwal et al., 2012; Harris, 2009; Levin & McEwan, 2001) was utilised to make sure that all actual costs for implementing the specific intervention were properly captured in the model. Additional forest plots with estimates of CER of the various interventions could then be plotted with their respective confidence intervals, to provide additional information useful to decision-makers.

5.5. Systematic review protocol

5.5.1 Inclusion criteria

"The key concern in any meta-analysis is that 'apples and apples' are compared and not 'apples and oranges'. Each individual study presents, in fact, a specific context, type of intervention, outcome being addressed and evaluation design. Yet meta-analytical studies need to be comparable conceptually, methodologically and statistically."

(Duvendack, Systematic Review Symposium, Bangladesh, 2012)

The studies that were included in the following systematic review met the following specifications:

a. ***Study designs*:** Rigorously performed experimental and quasi-experimental impact evaluations that utilised pre- and post-measures for both treatment and control groups. Sample sizes were relatively large to allow for enough statistical power of the results.
b. ***Participants***: Only studies conducted in South Africa's schooling sector, both primary and secondary phases (grade 1 to 12), post-1994. Both programme and control groups consisted of learners or schools from South Africa. This was to ensure a certain level of consistency with regard to context, social, economic, political, and systematic confounders.
c. ***Interventions***: Any type of interventions aimed at improving learning outcomes. These ranged from teachers-based, learner-based, materials and resources, management- or system-oriented, whole school development. See background section.
d. ***Outcomes***: Learner achievement, measured through test-scores or pass-rates. Particular focus was given to the domains of language, mathematics and science, but other learning results were also included.

5.5.2 Identification strategy

True to any systematic review, the study endeavoured to uncover all the published, and unpublished materials in order to reveal studies that would meet the above requirements. These included academic

papers, journal articles, research and evaluation reports from government, corporations and NGOs, consultants and any other grey literature or papers hidden in 'file drawers'.

Search terms and key words utilised included: *impact , *evaluation , *South Africa , *education , *effects , *learning, *achievement, *mathematics , *language, *programme, *outcome, *projects, *school, *teacher, *numeracy, *literacy – or combination of these.

In the literature search the following sources were consulted:

1. Databases and libraries of specialised agencies

Department of Basic Education (DBE)
Zenex Foundation[12]
Joint Education Trust (JET)
Human Sciences Research Council (HSRC)
Centre for Development Enterprise (CDE)
Education Research Agency (ERA), Stellenbosch University[13]
Campbell Collaboration (CC)
International Initiative for Impact Evaluation (3ie)

2. Academic databases and journals

Wits University library and electronic databases
Google scholar internet searches
South African Journal of Education
African Journal of Research in SMT Education
Evaluation and Programme Planning
International Journal of Evaluation Research

3. Collegial networks and informal communication with prominent evaluators and researchers in South Africa's education sector (Eric Schollar, Stephen Taylor, Cas Prinsloo, Brahm Fleisch, Jennifer Roberts, Martin Gustaffson, Anil Kanjee, Charles Simkins, Anne-Marrie Hattingh, Paul Hobden).

[12] Thanks to the review and cataloging recently done by Mouton et al. (2014), these were the database where most reports and publications where extracted from.
[13] Ibid.

4. Other South African reviews, meta-evaluations, primary studies and edited books including:

Bloch (2009). *The toxic mix: What's wrong with South Africa's schools and how to fix it*? Tafelberg: Cape Town.
Fleisch (2008). *Primary education in crisis*. Juta: Cape Town.
Mouton et al. (2013). *Zenex Review Project Final Report*. Johannesburg: ERA.
Roberts & Schollar (2006). *Meta-evaluation of interventions in math, science and language 1998-2006.* ESA.
Roberts & Schollar (2011). Meta-Evaluation of programmes and projects supported by the Zenex Foundation 2006-2011. ESA
Sayed, Kanjee, & Nkomo (2013). The search for quality education in post-apartheid South Africa. Johannesburg: HSRC Press.
Taylor, Flesich, & Schindler (2008). *Changes in Education Since 1994*. SA Presidency
Taylor, Muller, & Vinjevold (2003). *Getting schools working: Research and systematic school reform in South Africa*. Cape Town: Pearson Education, Maskew Miller Longman.
Taylor & Vinjevold (1999). *Getting learning right: Report of the President's education initiative research project*. Johannesburg: JET.

5. Previous international reviews and meta-analyses on education in developing countries:

Nye et al. (2006). *Approaches to parent involvement for improving the academic performance of elementary school age children.* Campbell Collaboration.
Ritter et al. (2006). *The effectiveness of volunteer tutoring programs: A systematic review.* Campbell Collaboration.
Zief et al. (2006). *Impacts of after-school programs on student outcomes*. Campbell Collaboration.

Hattie (2009). *Visible learning: Synthesis of 800 meta-analysis relating to achievement.* Routledge[14]
Petrosino et al. (2012). *Interventions in developing nations for improving primary and secondary school enrollment of children: a systematic review,* Campbell Collaboration
Baird et al. (2013). *Relative effectivenes of conditional and unconditional cash transfers for schooling outcomes in developing countries: A systematic review.* Campbell Collaboration.
Kremer, Brannen & Glennerster (2013). *The challenge of education and learning in the developing world*. *Science*, *340*(6130), 297–300.
Krishnaratne, White, & Carpenter, E. (2013). *Quality education for all children? What works in education in developing countries.* Working Paper (20). International Initiative for Impact Evaluation (3ie)
Conn (2014). *Identifying Effective Education Interventions in Sub-Saharan Africa: A meta-analysis of rigorous impact evaluations*. New York: Columbia University
Glewwe et al. (2014). *School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010.* Education Policy in Developing Countries, Chicago: University of Chicago Press
McEwan (2014). Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. *Review of Educational Research*
Murnane & Ganimian (2014). *Improving Educational Outcomes in Developing Countries: Lessons from Rigorous Impact Evaluations*, NBER Working Paper. N. 20284.

It is interesting to note that in a recent review conducted by Evans and Popova (2015) of six major international reviews in education published in 2013 and 2014, such studies reached strikingly different conclusions. Evans and Popova (2015) highlighted that only 3 out of the total 301 primary studies were to be found in all six of the systematic reviews. Though this was greatly due to the different inclusion

[14] Hattie's meta-analytical synthesis included many studies that were not necessarily experimental (had only single group pre and post results but no counterfactual).

criteria and intervention classifications that the authors chose, it still begs the question on how exhaustive and comprehensive were the searches conducted in these international reviews? It is thus crucial to ensure that the following local meta-analysis includes as much of the evidence stemming from the latest education evaluations conducted in South Africa.

5.5.3 Appraising quality and bias in the studies

Once all the studies that met the inclusion criteria had been identified and short-listed from the literature, one of the most complex subsequent exercises was the screening of individual evaluations on methodological rigour and design quality, to decide whether they should be included or not in the meta-analysis. The tension always lies between not leaving out any useful and important piece of evidence while also making sure that a few bad studies do not influence negatively the overall findings of the meta-analysis (Hombrados & Waddington, 2012). In deciding which studies are methodologically sound, careful assessments needed to be made on internal and external validity, construct measures and statistical errors.

In the medical field, meta-analyses restrict themselves almost entirely to rigorously implemented RCTs. In the social sciences and in education policy, as we have seen earlier, the field is dominated by quasi-experiments, which can present numerous different threats to validity and risk of biases. These however can be mitigated if treatment/control allocation rules are clear (Hansen et al., 2011), selection bias has been appropriately addressed, and programme implementation and causality are carefully controlled. Whether big or small, any bias, assumptions, errors or problems in a particular study need to be carefully examined and reported, so that these can be taken into account in the meta-analysis and used also as controls in the subsequent sensitivity and moderator analysis.

More than 50 different threats to validity exist in both experimental and quasi-experimental studies, however in the current meta-analysis on the interventions in South Africa's education sector, the following aspects have been given special attention in the evaluation quality appraisal process:

1. *Selection bias* – if the counterfactual is not strong or if programme and control group are practically and significantly different. The experiment therefore becomes compromised from the start.

2. *Implementation bias* – if the programme was not implemented correctly according to plan, or if the two groups received essentially different treatments from different service providers. Implementation bias can also include spillover between treatment and control group and contamination of the samples by other interventions.

3. *Data and testing bias* – if a large portion of the data is unreliable, inaccurate or missing. Large amount of attrition occurring or no-show of participants between pre- and post-tests. This can also include the presence of 'ceiling effects' or 'floor effect' in the data gathering process.

4. *Evaluator bias* – if the authors of the evaluation had a conflict of interest (political or financial), were directly involved in the design or implementation of the programme, and might have a vested interest in the results of the study. This comprises the independence of the study.

5. *Motivational bias* – if there is a strong presence of Hawthorne (Stand, 2000) or John Henry effect (Saretsky, 1972), where the participants of either the treatment and control group behave differently as they are conscious of being observed, and have a vested interest in the results of the evaluation.

In education evaluation there is often a tension between trying to avoid selection biases and contamination biases. To allow for similar characteristics in programme and control groups, schools and learners are often matched based on proximity of location, however this can also increase the likelihood of control participants taking advantage of the intervention when they are not supposed to. These challenges have been illustrated also by Schollar (2014) in his RCT of the PMRP in Limpopo province. These highlight again the many practical and ethical complications of conducting experimental evaluations in real-life scenarios, such as in the education policy field.

Many tools and instruments exist to assess risk of bias developed by Cochrane Collaboration, AHRQ, NICE, Wells, DIAD, Maryland and Duvendack (2011). Rating studies however is generally very controversial, subjective and discouraged as a source of reliable quality assessment (Juni et al., 2001; Wilson et al., 2011). Assessing the quality of studies ultimately remains a judgement call based on expertise of the reviewer both in statistics and the content matter.

Having acknowledged all the limitations, this systematic review uses an adapted version of the Cochrane risk of bias tool (Sterne, Higgins, & Reeves, 2013). After a careful assessment by the review team, each study was rated against the 5 major bias areas discussed above. A score from (1) to (5) was assigned to each study signalling the degree of bias and problems detected (Table 24):

Table 24 *Risk of Bias Rating*

Rating	Level of Bias
1	Minimal
2	Low
3	Moderate
4	High
5	Critical

Practically all of the studies that were included in this meta-analysis had pre- and post-scores for both treatment and control groups, allowing them all to integrate the DID method in the evaluation design (see section 2.13). The consistent use of double difference method reduced some of the biases by combining the time variable (before and after) with the counterfactual variable (participants and non-participants). Any inherent differences between programme and control groups because of observed or unobserved differences was significantly reduced, assuming that these differences remained constant over time (Blum et al., 2010). This is the case with most studies and interventions included in this meta-analysis as the learner population was affected by the same systemic and external influences in the same country during the same period of time.

5.5.4 Data extraction and follow-up

As discussed earlier, the effect size used in this meta-analysis is Hedge's *g*, which is an enhanced version of the SMD and Cohen's *d*. To compute this effect size what is essentially required is mean scores, standard deviations and the sample sizes for both treatment and control groups in both pre- and post-tests (as most of our studies were DIDs). These statistics also allow for the calculation of the confidence intervals and standard errors for each effect size. The statistical data required from each primary impact study can thus be summarised through Table 27.

Table 27. *Summary Statistics Gathered for Studies in the Meta- Analysis*

Programme Group						Control Group					
Pre-test			Post-test			Pre-test			Post-test		
Mean	Std	N	Mean	Std	N	Mean	Std	N	Mean	Std	N

In many studies the sample sizes of the participants would not be the same in the pre- and post-- tests, because of learner attrition, thus posing a threat to validity in the implementation of the experiment. Other times because pre- and post-tests were not done with the same cohort of participants, but rather longitudinally with the same grade level. In such cases the reviewers needed to decide to either use the average means or to use only the values from the individuals that appeared in both pre- and post-tests. These situations are explained when discussing the merits of each individual evaluation design (see section 5.2).

The 12 statistics (Table 27) were reported in their entirety in only some of the studies that were reviewed. In the most of the cases the standard deviations were not reported, therefore the reviewer had to contact the original authors to request the missing data. Other times standard deviations would be re-calculated manually by going through the raw data provided by the author of the primary study. Following up with the original authors provided the reviewer also an opportunity to seek further clarification on the evaluations, the interventions and other important details of the studies. Except for the Dinaledi evaluation, which was conducted by World Bank economists based in Washington DC, following up with the original authors on clarifications and data requests was relatively easy, as they were all based in South Africa and easily accessible through informal e-mail, telephone communication and face-to-face discussion.

The data collection and data cleaning, the risk of bias rating, and the coding of the various studies (discussed later in the section 5.7.3) was conducted by a team of reviewers (the primary researcher and two research assistant) to allow for double-checking of the accuracy of all the data throughout the process. In some complex situations, the reviewers had to debate how to handle a particular case and take decision on which data to use. These special cases will be explained in the next section. Once the required statistics were checked and extracted from all the studies, effect sizes and confidence intervals were calculated utilising CMA statistical software package.

For a couple of studies, numerous attempts were made to contact the original authors of the papers to seek the missing statistics. However response was not received and these studies had to be unfortunately excluded from the meta-analysis:

- Niewoudt, S., Hieuwoudt, H., Monteith, J. (2007). Influence of a video class system on learners' study and learning strategies and their achievement in mathematics. *African Journal of Research in SMT Education*, *11*(1), 29-35.
- King, L. (2001). Assessing the effect of an instructional intervention on the geometric understanding of learners in South Africna Primary Schools. Paper presented at AARE Annual Conference, Freemantle, 2001.

For the study cited below, the author advised the review team not to include the evaluation in the meta-analysis as the final results of the impact study had not be processed yet:

- Prinsloo, CH, Rogers, S (2014), Hope for rebuilding language foundations: Evaluation report of the siyaJabula siyaKhula's Learner Regeneration Project, Human Sciences Research Council: Pretoria.

Similarly, the reviewers were aware of another major education RCT being conducted in North West province by Wits University, HSRC and DBE. However, at the time of preparing this meta-analysis, the results of the North West study was not yet available. Nonetheless these two impact evaluations will be included in future updates of this South African systematic review.

5.6. Qualitative results of the systematic review

5.6.1 Overview of studies and interventions included in the review

At the end of the systematic literature search, the eligibility screening and data extraction phases, the review yielded a grand total of 28 studies, 31 interventions (Table 28) and 127 effect sizes, based on impact measured against different learning outcomes (language, mathematics, science, etc.) or within different sub-groups (i.e. classes, cohorts, years, intervention dosage). Similar to McEwan's (2014) review of education interventions in developing countries, most experimental evaluations included in this meta-analysis had 'multiple treatment arms'. Each primary study in fact reported anything from 1 to 22 effect sizes. Some studies assessed more than one intervention (i.e., Schollar, 2001; Besharati, 2014) and two studies were found for the same interventions (Fleisch et al., 2011; Schollar, 2015).

Table 28. *Studies and Interventions Included in the Systematic Review*

Evaluation	Intervention	Learning Outcome	Other sub-groups
Besharati (2014)	Proximity to Mine Amplats programme Radical M&S winter schools	Maths Physical Science Language Overal NSC passes	30% Passes 50% Passes
Bohmer (2014)	Numeric CAL after-school programme	Maths	
Botes & Mji (2010)	Language companion	Maths	
de Chaisemartin (2010)	Cape Town Leadership Institute	English Maths	Normal dosage High dosage Grade 3 Grade 6
Feedback R&A (2008)	Maths Centre	Maths	2006 Results 2007 Results
Fleisch, Taylor, et al. (2015)	RCUP	Literacy	ANA test JET test
Fleisch & Schoer (2014)	GPLMS	Literacy Numeracy	
Fleisch et al. (2011)	B2B Workbooks	Maths	Normal dosage High dosage
Gustafsson et al. (2013)	Change in Provincial Boundaries	Maths Scores Maths HG Passes	2011 results 2012 results
Hattingh (2003)	Problem-Based Learning	Science	
Hobden & Hobden (2009)	RUMEP	Maths	Basic Skills 2007 Basic Skills 2008 Content
JET (2006)	Mother Tongue Literacy Program	Language	
JET (2007)	Mindset Network	Maths Science	Grade 10 Grade 11 Cohort I
Louw et al. (2008)	Khanya ICT	Maths	
Padayachee et al. (2011)	DVD Blend	Maths	
Prinsloo & Kanjee (2005)	Quality of Learning Programme	English Maths NSC passes	Grade 9 Grade 11 NSC English HG NSC Maths SG Overall NSC passes
Prinsloo (2009)	Plus Time	English Maths	English L1 English FAL
RIEP (2008)	MCPT	Literacy Numeracy	isiXhosa isiZulu

Table 28. *Studies and Interventions Included in the Systematic Review (continued)*

Evaluation	Intervention	Learning Outcome	Other sub-groups
Roberts and JET (2006)	COUNT Numeracy programme in Uitenhage	Numeracy	Grade 3 (2004) Grade 3 (2005)
SAIDE (2007)	COUNT Family Math	Math	Grade 2 Grade 5
Schollar (1999)	MPSI INSET	English Maths	Grade 4 Grade 5 Grade 6 Cohort I Cohort II
Schollar (2001)	Read Kei Komga Read Transkei	Read Write	Grade 3 Grade 4 Grade 5 Grade 6 Grade 7 Cohort I Cohort II
Schollar (2002)	EQUIP	English Maths	Grade 6 Grade 7 Cohort I Cohort II
Schollar (2005)	Learning for Living	Read Write Maths	Grade 5 Grade 7 Cohort I
Schollar (2015)	B2B Workbooks	Maths Grade 4 Maths Grade 6	Normal Dosage High Dosage 5 years later
Schollar & Mouton (2014)	Epoch & Optima Maths Challenge Programme	Maths	30% pass rate 50% pass rates 60% pass rates weighted average score
Taylor et al. (2015)	MTG Study Guides	Geography Accounting Economics Life Sciences	
World Bank DIME (2010) or Blum et al.	Dinaledi	Physical Sciences	Standard Grade Higher Grade

Once again it is interesting to note that of the 301 impact studies on education interventions in low and middle income countries, captured in the six international reviews discussed by Evans and Popova (2015), only one of the South African studies listed above (Louw et al., 2008) was found in only two of the international reviews (Conn, 2014; Glewwe & al, 2014). This points to a serious publication bias present in most international systematic reviews, that give more prominence to academic studies conducted by Northern scholars. The international education reviews conducted in 2013 and 2014 appear to have overlooked a large portion of the knowledge produced locally in South Africa, which often is published as evaluation reports by governmental, research and private education agencies active in the sector (see more in next section).

5.6.2 Qualitative assessment of the evaluations included in the review

Before embarking in the comparative meta-analysis of the effect sizes of the various interventions, this study integrates also a qualitative component by providing a narrative review of the various South African education programmes/policies and their evaluations. This section thus provides a detailed description of each of the education interventions included in this systematic review as well as a commentary on the quality, strength, weaknesses and limitations of the studies that were conducted for each of the interventions. For each of the primary studies the potential biases and threats to validity are discussed, and methodological quality of the evaluation is rated based on the scale discussed in section 5.5.3.

Because of the practical constraints discussed earlier, full-fledged RCTs in South Africa's education sector have been few; however, there have been instances where evaluators have used natural experiments (Gustafsson et al., 2013; Taylor, 2012) and RDD (Besharati, 2014; Fleisch & Schoer, 2014), to achieve internal validity in education impact studies. As explained earlier the vast majority of the evaluations included in this meta-analysis have used the combination of DID with matching techniques of different forms and rigour, thus requiring close scrutiny with regard to assumptions, biases and threats to validity. The way sampling was done also plays a large role in the confidence of the effect size estimates.

Below is a discussion on each of the interventions and evaluations that were included in this systematic review. Some of the studies had very high risk of bias and other methodological problems that precluded them from being included in the subsequent quantitative meta-analysis (see next section). For analytical purposes the interventions have been grouped into (a) government programmes and policies; (b) private sector projects; (c) academic experiments.

a) Government programmes and policies

In South Africa, National and Provincial Departments have different mandates and functions in the education system. Both entities have a certain degree of freedom to pilot different policies and programmes with the schools under their jurisdiction. Below are a few of such initiatives by various education authorities implemented in different provinces of South Africa or at national level. These policies were typically evaluated by researchers and advisors linked to the government, or by expert teams funded by foreign donors (World Bank, DFID, USAID, etc.), who have supported the South African government in addressing the country's education challenges.

Intervention 1: Khanya Technology in Education Project

The Khanya project was an initiative of the Western Cape Department of Education (WCDE) which aimed at improving the delivery of mathematics curriculum through the use of an ICT-based system. This would compensate for the low curriculum coverage due to poor teacher capacity. Schools from the province that were considered both poor but also well-managed were provided with computers, MasterMaths (MM) software package and a MM tutor who would administer the programme and the assessments. The software would assist in covering the official mathematics curriculum from grade 7 to 12. From 2001 to 2006 more than 595 schools benefitted from the initiative.

Evaluation 1: Johann Louw, Johan Muller, Colin Tredoux (2008), University of Cape Town

The authors of the study utilised a premature (conducted too early) outcome evaluation of the Khanya project to conduct an experiment on the impact of the initiative on learner performance in grade 12 mathematics. NSC exams were used as post-test while the pre-test consisted of grade 11 self-administered school assessments; therefore, test instruments were not consistent across schools and across time.

Although the study endeavoured to be an RCT, it finally resembled an ex-post quasi-experiment which utilised matching methods. Programme and control groups were created by randomly sampling learners from five intervention and five control schools matched on similar demographics, social-economic parameters, and taking into account Khanya implementation levels. The scores of the two groups at baseline were significantly different, in large part due to an outlier school in the control group that performed particularly poorly. Learners were drawn from too small a sample (five control and five treatment schools); therefore, both the internal and external validity of the study is very weak and therefore the standard error and confidence interval of the results are also very high.

The study concluded that there was no clear indication on the effectiveness of the Khanya project on learning outcomes. Nevertheless, regression analysis was also undertaken in the study, that showed that improved mathematics performance was correlated to learners spending more time on the software, as well as other classroom teaching and social-economic predictors. *Risk of Bias: 4*

Intervention 2: Cape Town Leadership Institute

The Cape Town Leadership Institute (CTLI) is an in-service teacher training center set up by the Western Cape Department of Education which has been functioning since 2002. Teachers and school managers are regularly trained through block-release courses with the aim of raising the quality of teaching practices and therefore improving learning outcomes in public schools of the Western Cape.

Evaluation 2: Talia de Chaisemartin (2010), JET Education Services

In 2010 JET Education Services was asked to conduct an evaluation of the CTLI, which included both a process/implementation evaluation as well as impact evaluation, which looked at the quality of the courses, the transfer of knowledge, the change in teaching practices and finally the improvement in learner achievement. Learner results were assessed utilising the provincial systemic tests for both language as well as mathematics. Average school results were used as units of analysis. Grade3 results in 2002 and 2008 functioned as pre- and post-test for the foundation phase, while changes in grade-6 between 2005 to 2009 were used for assessing impact on intermediate phase. Schools were categorised by the ones who had not benefitted from any CTLI training (control group), the ones who had 1-4 teachers trained in the time-period (standard programme group), and the schools who had 5 or more teachers trained (high dosage programme group). The evaluation found that there was a significantly higher improvement in learning results for the schools that had benefitted from CTLI trainings, particularly in the area of language. Effects were also larger in the schools that sent more teachers to participate in the CTLI programmes.

The author was very collaborative in providing the original dataset to the reviewers to calculate effect sizes for the meta-analysis. The evaluation however appears to suffer from high selection bias, as even the evaluator herself discovered that, the majority of the schools participating in the CTLI were already high-performing, coming from urban areas and from privileged quintiles (de Chaisemartin, 2010). This was not only contrary to the original mission of the CTLI, but also did not allow for a proper counterfactual. The schools that did not participate in the programme were generally the weaker, poorer and more rural schools. There were also some data challenges which made it difficult to confidently link the learner results to the presence of a specific teacher trained by CTLI, as there was a lot of course drop-outs and teacher mobility inside and between schools. The time-frame between pre- and post- tests was also very long (4 - 6 years) which made possible for other factors to have an influence on the learning outcomes, beyond just the CTLI. *Risk of Bias: 4*

Intervention 3: Mpumalanga Primary Schools Initiative (MPSI)

The MPSI was a programme funded by DFID aimed at improving learning outcomes in primary schools in the Mpumalanga province in the late 1990s. The MPSI also had a prominent teacher development model which consisted of in-service training, classroom monitoring and feedback, lesson plans, and a wide range of LTSMs. The theory underlining the programme was that appropriate resources, knowledge, behavioural and attitudinal changes in teachers, would improve classroom practices and hence learner achievement.

Evaluation 3: Eric Schollar (1999), ESA

Eric Schollar was invited to conduct the summative evaluation of the INSET component of the MPSI. Pre- and post-ESA tests for language and mathematics were administered for grades 4, 5, 6 and for two cohorts of same learners across years. Programme and control groups were matched based on simple demographic characteristics. The overall results did not show major effects on learner achievement, except a bit for the results in grade 6, specifically for language. The author warned that the impact assessment was done only 18 months after the programme had started therefore on one hand there might have not been enough time for the intervention to mature, but on the other hand there was also high potential for Hawthorne effects.

The author had also warned that the programme group results might have been skewed by the high performance of one outlier school, Koornfontein, which had much better facilities, more trained teachers, strong management and a wide range of LTSM, not comparable to the other schools in the sample. Having had received the complete raw data from the author, the reviewer calculated the 10 effect sizes from this study excluding the results from the learners in Koornfontein school. *Risk of Bias: 3*

Intervention 4: Mind the Gap (MTG) Study Guides

In 2012, the National Department of Basic Education (DBE) launched a series of study guides to assist learners to better prepare for the NSC exams. The Mind the Gap (MTG) study guides were designed to be supplementary self-study material that would assist learners in weakly performing schools where curriculum coverage may have been incomplete and learners lacked exposure to the content being tested in the NSC. The Study Guides were available on the DBE website but to assist poor learners who lacked connectivity the Department distributed hard copies to schools in certain underperforming districts around the country. In 2012, MTG study guides for Accounting, Economics, Geography and Life Sciences were produced for the province of Mpumalanga; however, the available budget (funds for transport costs and the number of guides already printed) dictated that only a limited number of schools could benefit from the LTSM.

Evaluation 4: Stephen Taylor and Patricia Watson (2015), University of Stellenbosch

The researchers that led this study used the limitations in the distribution of the study guides as an opportunity to work with the DBE to conduct an RCT to assess the impact of the programme on grade-12 NSC results. Once the sampling frame was defined, a computerised lottery was used to select 79 treatment schools who would receive the study guides, leaving 239 schools to function as control group. Considering the allocation was done through proper randomisation we can assume a strong and legitimate counterfactual.

Due to printing delays, study guides reached all recipient schools 1 to 2 month before the examination, which coincidently was the period when these materials would be most utilised. Even though the guides were available on the DBE website, spillover to learners in control groups was minimal as most of them lacked internet connectivity and awareness on such material. Implementation of the experiment was thus also well controlled. Learner outcomes were observed by reviewing NSC results available in DBE databases, thus avoiding intrusion in the population or risk of Hawthorne or John Henry effects. For all intents and purposes the RCT was text-book executed with a very high degree of technical rigour.

Impact of the programme was measured by comparing NSC results in 2011 (pre-test) and 2012 (post-test), after the study guides had been distributed. The results of the experiment concluded that two of the study guides (Accounting and Economics) did not show a statistically significant difference between the treatment and control group, but the guides for Life Science and Geography did have a significant effect on the NSC scores, although learning gains were still relatively small. The author went on to explore the potential reasons and contextual factors affecting the heterogeneity in the findings of the study. As part of the paper, Taylor also conducted OLS regressions, a cost-benefit analysis, and a simulation of taking the intervention to scale.

The quality of this study is one of the best within this meta-analysis. Analysis was done at learner level and sample sizes were extremely big which helped keep standard error to a minimum. The author provided the reviewers with the basic statistics required to calculate the effect sizes for each of the four subject study guides. To allow for comparison of these results with the learning outcomes from the majority of the other studies in this meta-analysis, accounting was grouped under 'mathematics' and economics and life sciences under 'sciences'. *Risk of Bias: 1*

Intervention 5: Gauteng Primary Language and Mathematics Strategy (GPLMS)

The GPLMS is an intervention of the Gauteng Department of Education that started in 2010 aimed at improving language and mathematics abilities of learners at the foundation phase. In essence it provides teachers with highly scripted daily lesson plans, high-quality LTSMs and a programme of regular monitoring and coaching. The GPLMS is targeted at historically disadvantaged and underperforming schools of the province.

Evaluation 5: Brahm Fleisch and Volker Schoer, 2014

Grade 3 learner test scores were analysed, utilising the 2008 Systemic Evaluations as the baseline and the 2011 and 2012 Annual National Assessments (ANA) as the post-test. This is the only study that utilised a rigorous RDD to create the counterfactual, using different sample specifications closer or farther to the cut-off assignment threshold. It also presented additional fixed effects modeling. Units of analysis were schools, therefore the grade 3 learners of each year of assessment were different. Fleisch and Schoer reported that their results are inconclusive as they suspect a potential test instrument effect which might allow low-performing schools to gain more than high-performing schools. There does appear to be a LATE at the assignment cut-off, but it is not clear if the effect is generalisable to the rest of the population. The authors also warn against possible Hawthorne effect and potential 'gaming', as the ANA tests are marked by the same teachers under observation.

It also must be noted that one of the authors of the study is a senior advisor to the Gauteng Department of Education and was closely involved in the design of the intervention, therefore might have a bias with regard to the results of the evaluation. The reviewer was sceptical about the use of different test instruments (SE and ANA) in the pre- and post-tests. To also allow more time for the intervention to mature, the reviewer requested from the authors the results also from 2013 ANA so that these can be used as post-test while the 2011 ANA could be used as pre-test. The original paper reported only literacy results, but the reviewer requested the results also for the numeracy test scores. The effect sizes reported in this meta-analysis therefore reflect these additional adjustments made by the reviewer to the updated data received from the authors. *Risk of Bias: 2*

Intervention 6: Change in Provincial Administration

In 2005, seven of the South African provinces saw their boundaries change to ensure that all municipalities fell into just one province. This affected a total of 710 schools (158 high schools) who switched from one provincial education administration to another. Provincial authorities have significant powers and influence on education policy and on the functioning of the schools under their jurisdiction. They are responsible for managing resourcing, in-service teacher training, and local accountability systems. There is a common perception in South Africa that certain provincial administrations (such as Gauteng and Western Cape) function better than others.

Evaluation 6: Martin Gustafsson and Stephen Taylor (2013) , University of Stellenbosch

The authors of this paper decided to utilise the change in provincial administrative boundaries, to conduct a natural experiment to test the hypothesis whether provincial administration does have an impact on learning outcomes. Different indicators from school results of 12-grade NSC exams were observed using various models – value-added school production, DID, spatial analysis, fixed effects panel data regressions. The particular focus of the paper became the switch of the 29 schools from North West province to Gauteng, which occurred between 2007 and 2008. The study found that the mathematics NSC average scores and high-end passes of the switching schools did indeed benefit from a significant increase; however, this might have also been masked by a large fluctuation over the years of average NSC results experienced by both provinces. Though not fully conclusive, the paper makes a modest contribution to empirically confirming the public perception that provincial administration does have an impact on the overall education results.

The study was not originally reported as a classical quasi-experiment, therefore the reviewer requested the authors to provide the required statistical data and to decide which of the many sample specifications presented in the paper was going to be used as the control group for the natural experiment. The control group chosen was 'all other secondary schools in North-West province' which provided a fairly weak counterfactual to the 29 switching schools at the border, which could have been quite different from the rest of the schools in the province. The reviewer replicated the quasi-experiment utilising both 2012 and 2011 as the post-test scores and did indeed find and confirm that there was significant difference in the results from one year to the other, probably due to major surges in NSC results of both Gauteng and North-West. Effect sizes of both years are reported in the meta-analysis. The time-frame for the experiment 2005 to 2012 was also fairly long with many other factors that could have influenced the learning outcomes aside from provincial boundary change. Furthermore the change of provincial administrative responsibilities coincided with a change in 2007/2008 of the NSC examination instrument, which could have also caused some blurring of the pre- and post- test results as these were performed essentially with different instruments. *Risk of Bias: 4*

Intervention 7: Dinaledi Schools

Dinaledi (meaning 'Stars') was an initiative of the National Department of Education launched in 2001 as part of the strategy to improve mathematics, science and technology in public high schools. More NSC passes in these subject were required in order to address the shortage of technically skilled professionals in South Africa's labour

markets. The programme aimed to double the number of mathematics and science passes at the NSC to 50,000 by 2008. Schools were selected to be 'Dinaledi' based on a combination of needs and minimum capacity criteria. Dinaledi schools were typically poor, rural and township schools which however have good mathematics and science throughputs and were considered centers of excellence. Dinaledi schools would receive special attention and be given additional teachers, in-service training and LTSM (such as textbooks, calculators, computers, science kits, audio-visual materials), to strengthen the teaching and learning of mathematics, science and technology. Initially starting with 102 schools, the programme gradually expanded in 2006 to reach over 513 schools across the country. Dinaledi is a flagship programme of the Department of Education, but also received over the years significant support from various private sector funders, thanks to the 'adopt a school' initiative. Corporations that supported Dinaledi schools over the years include ABSA, Anglo American, IDC, Nokia/Siemens, Old Mutual, Epoch & Optima, Standard Bank, Shell, Transnet, Telkom, WBHO, Woolworths, and Zenex.

Evaluation 7: Jurgen Blum, Nandini Krishnan, Arianna Legovini (2010), World Bank

In 2010 a team of economist from the World Bank assisted the DBE to conduct a rigorous impact assessment of the Dinaledi programme in relation to NSC enrolment and pass rates in mathematics and science. By combining large amounts of administrative, infrastructure and test score panel data from various DBE databases (EMIS, NEIMS, PERSAL and NSC), the evaluators were able to conduct a retrospective impact assessment. The Dinaledi evaluation combined PSM (Heckman et al., 1998) with DID method to estimate the effects of the programme. Through sophisticated statistical procedures (ie. PSM), a sample of 350 Dinaledi schools was matched to a similar number of control schools based on school demographic characteristics and previous performance on the NSC. Data points from 2004 were used as the pre-test and 2007 as post-test. The evaluation concluded that Dinaledi did have a significant and substantial impact (ATE) on the enrolments as well as passes in physical sciences at NSC of the participating schools. The findings showed that in Dinaledi Schools, 57% of learners passed mathematics and 59% passed sciences, which was way above national averages. The study also explored heterogeneity and found that Dinaledi had the greatest effect in the schools under former Bantustan administration, where results increased sevenfold. The mathematics NSC data, especially around 2004, unfortunately was not reliable and therefore could not yield to conclusive findings.

Most of the results of this study were reported through fixed effect regression models and not in the classical experimental fashion. The reviewers had to thus follow-up with the DBE and with the World Bank to seek clarification on the statistics that were not reported in order to calculate the effect sizes required for the meta-analysis. After numerous attempts to contact the World Bank, the reviewers had to finally work off the original datasets and replicate the entire Dinaledi impact evaluation, using the same techniques and procedures used in the World Bank study in order to confirm the missing statistics. The main difference between the World Bank evaluation and the reviewer's replication was that the first utilised STATA software package while the second utilised SPSS, nevertheless the final results were very similar. For the sake of consistency across the meta-analysis the science learning results extracted from the World Bank study were the ones reported as percentages (ie. pass-rates in science SG and science HG) rather than the ones reported in absolute numbers.

Although the World Bank went through some great length to implement very sophisticated impact evaluation techniques, the study presented a number of limitations with regard to the DBE data (not always present or reliable). Also there might have been some selection bias and pre-existing differences between the control schools and the Dinaledi schools, which were per se selected into the programme because they were already among the best high schools in the country and thus likely to perform better than others during the experiment. There was also a large degree of inconsistency in the Dinaledi implementation, as different Dinaledi schools received different treatments by different provincial departments of education and different private companies during the course of the years. *Risk of Bias: 3*

Intervention 8: Reading Catch Up Programme (RCUP)

The Reading Catch Up Programme (RCUP) was designed as a component of the Gauteng Primary Language and Mathematics Strategy (GPLMS) (see above), implemented by Class Act as a remedial programme in English language for intermediate school phases (grade 4 to 6). After a positive evaluation of the programme in Gauteng (Hellman, 2012), a research team funded by Zenex Foundation decided to conduct a rigorous RCT of the RCUP in Pinetown district in collaboration with the Kwa-Zulu Natal Provincial Department of Education. The experiment was conducted in the first term of 2014. The intervention consisted of a 11-week programme of on-site teacher coaching (through training, monitoring and support) accompanied by the provision of LTSMs such as scripted lessons plans, learner exercise and assessment books, and other reading material. RCUP was implemented in both rural and urban schools, however the implementation was not consistent with different degrees of compliance with the teaching approaches and utilisation of LTSMs. The programme was very ambitious and schools managed to cover only 63-74% of the RCUP curriculum.

Evaluation 8: Brahm Fleisch, Stephen Taylor, Volker Schoer, Thabo Mabogoane (2015)

In order to address the limitations of the Hellman (2012) study (pre-post evaluation administered by the implementing agency), the research team felt it was important to test the effectiveness of the RCUP using a counterfactual evaluation.

Through a rigorously performed RCT and a careful process of cluster sampling, 40 treated and 57 comparator schools were selected in the Pinewood district, which met the criteria of being English-medium public schools in quintile 1-4 that performed below 55% in the ANA 2013. Schools generally had only 1 language teacher, teaching between 15-120 grade 4 learners. Baseline results in both control and treated schools were very similar, thus confirming the strong counterfactual developed in this study.

Pre- and post- literacy assessment were conducted utilising test instruments designed by JET Educational Services, as well as examining the grade 4 ANA results in 2013 and 2014. While utilising the ANA results was more problematic, the JET assessments allowed following the achievement of 2543 treated and control learners over the year, with a very small attrition rates (less than 8%).

As both treatment and control groups had impressive gains from the pre- and post-test, the experiment proved that the RCUP had very small impact on language outcomes of recipient schools. This could have been in

part due to normal developmental processes or to Hawthorne, John Henry, and floor effects. Nonetheless, the study underlined the importance of counterfactual evaluations when assessing effectiveness of education interventions.

Heterogeneous treatment effects were also explored including the results in various grades, components of the test, based on gender and initial competencies of learners, and quality of coaches - which proved to be a strong factor influencing the results of the programme. The evaluation highlighted the still large reading ability backlogs present in grade 4 learners of primary schools in South Africa. *Risk of Bias: 2*

b) Private sector funded projects

As discussed in the background section, South Africa has provided a unique case study of private sector engagement in education development (see also section 3.7.3). Domestic and foreign companies have provided substantial contributions to the country's education sector, through their CSI, philanthropic foundations and through numerous types of public-private partnerships. Much of these private sector initiatives would be implemented through specialised NGOs and service providers, but always also in close collaboration with national and provincial departments of education. The meta-analysis covers some of these private sector programmes implemented by some of the major institutions involved in the sector, such as:

- **READ Education Trust,** a prominent education NGO active in South Africa since 1979 dedicated to literacy development, through provision of materials and educator training.
- **The Joint Education Trust,** was founded in the early 1990s as a partnership between business, unions and black political organisation in order to address the challenge of restructuring the country's dire education sector. Since 1992, JET has spent over R1 billion in educational programming and policy, research and evaluation, continuously testing and refining school development models.
- **The Business Trust,** one of the largest public-private partnerships of South Africa, which involved cabinet ministers and some of the top business leaders on its board. The Business Trust operated between 1999 and 2011 and managed a total fund of R 1,8 billion, a substantial portion of which went for primary and secondary school development projects (see LFL and QLP).
- **National Business Initiative,** a voluntary coalition of corporations dedicated to promoting sustainable development. The NBI has been involved in the field of education since 1995, most prominently through the EQUIP programme.
- **Shuttleworth Foundation,** founded by South African space entrepreneur and philanthropist, to promote mathematics, science and technology education. It is noteworthy for its support of innovation and piloting new initiatives in the arena of education.
- **Zenex Foundation,** a non-profit grant-making agency dedicated to the improvement of language, mathematics and science outcomes in previously disadvantaged schools. Zenex Foundation has been a leader in South Africa in evidence-based research and programming in the field of education. It works through numerous implementing partners and a wide cadre of experts and evaluators.
- **Anglo American,** mining giant and one of the largest corporations (in terms of annual financial turnover and

number of employees), hence also the biggest social investor in South Africa. The various Anglo American business units (Anglo Platinum, Kumba Iron Ore, Anglo Coal, De Beers, etc.) as well as its Chairman's Fund (AACF) have been jointly spending nearly R700 million each year in social investments, with a large portion going to education.

Intervention 9: READ Primary School Programmes in the Eastern Cape

Between 1995 and 1998 the READ Educational Trust implemented two programmes in the Eastern Cape province aimed at improving English language ability of primary school learners. The Eastern Cape rural setting provided possibly one of the poorest social-economic context and some of the weakest education outcomes of the country. The READ model consisted in the provision of specialised books and LTSM, accompanied by teacher training and monitoring.

Intervention 9.1: READ Transkei

One of the above READ projects was implemented in 35 schools of the former Transkei homeland.

Intervention 9.2: READ Kei Komga

The other READ project was implemented in 37 schools of Kei Komga (former Ciskei homeland).

Evaluation 9: Eric Schollar (2001), IJER

The two READ interventions discussed above were evaluated both qualitatively and quantitatively by two agencies, CAST and ESA, who developed the learner language assessments. Through a pre- and post- quasi-experimental design, control and programme samples were developed through simple matching of school characteristics. Kei Komga intervention was evaluated with regard to reading and writing outcomes in grade 3, 5 and 7, while Transkei project looked at reading results for grade 5, 6 and 7. Both case studies also utilised cohort analysis where same learners were followed throughout the various grades. In both cases the assessment yielded extremely high results in favour of an impact produced by the programme on participating learners. This could be partially due to the unique historical, geographic and social-economic context in which the project operated.

When closely reviewing this study, a number of methodological problems also transpired. The Schollar journal article was based on data collected by CAST during the pre-test and ERA during the post-test. CAST folded up before the end of the project, so the reviewer had access only to the ERA original datasets. Standard deviations, which were not reported in the paper, could be re-calculated from the ERA post-test raw data, but for the pre-test these had to be only guessed or assumed to be similar to the post-test. Critical data was therefore missing, and sometimes the data reported in the paper did not match the data in the original dataset. There was also one major anomaly in the results which the reviewers decided to take out from the meta-analysis. This study was one of the oldest quasi-experimental studies implemented in South Africa's education sector and thus presented some major data problems, high risk of bias, sketchy results and thus if integrated in this meta-analysis needs to be analysed with extreme caution. *Risk of Bias: 4*

Intervention 10: Learning for Living (LfL)

Learning for Living was a major whole school development programme which ran between 2000 and 2004 in over 900 primary schools in all nine provinces of South Africa. It was funded by the Business Trust and implemented by READ. Its aim was to improve the quality of primary education, by tackling the critical area of literacy and language. This was done through three major strategies, namely a) in-service training for classroom teachers and school managers, b) provision of a range of books and LTSMs, c) provision of classroom monitoring and feedback. The whole school development project was implemented through a network of READ offices which would work very closely with the national and provincial education authorities.

Evaluation 10: Eric Schollar (2005), ESA

The project received a solid evaluation by Eric Schollar who conducted both a qualitative as well as quantitative impact assessment through a quasi-experimental design. The evaluation involved 80 principals, 152 teachers and 2678 learners divided in three-yearly cohorts examined separately. Learner reading, writing and mathematical skills were tested through ESA-designed test instruments administered to 50 project and 30 control groups at baseline (2000), mid-term (2002) and at the end of the intervention (2004). The evaluation utilised both a longitudinal cohort-based approach (following the progress of the same learners from one grade to the other) as well as grade-level testing (thus same school but different learners in different years) to measure progress on learning outcomes. Treatment and control groups were matched based on simple observable demographic and social-economic characteristics. All tests proved unmistakably a significant increase in learner achievement in project schools, particularly in the area of reading. The first cohort, who had been with the program for the longest time, provided the most reliable findings on the impact of the project. The evaluation also had a cost-per-capita assessment component that resulted R175 being spent on every learner.

The author provided the reviewers with the original comprehensive datasets from which eight effect sizes could be extracted for this study. At times the numbers presented in the report did not match the numbers available in the dataset. In such cases the reviewer chose to use the later. For the cohort-based studies a lot of attrition of participants was noted (up to 40%), however cohort 1 (grade 3 to 7) provided fairly reliable results. *Risk of Bias: 3*

Intervention 11: Quality of Learning Project (QLP)

The Quality of Learning Project (QLP) was a major whole school development programme, carefully designed and grounded in theory and past experience. It was funded by the Business Trust and implemented by a consortium of 10 NGOs led by the Joint Education Trust (JET), guided by the National Departments of Education. It was implemented in 524 high schools in all nine provinces of South Africa. Its aim was to improve learner achievement in secondary schools, with particular focus on language and mathematics outcomes as the foundation of all learning. The QLP model included a mix of interventions aimed at improving classroom teaching (through INSET and LTSMs), school management and governance, and district development and support.

***Evaluation 11: Cas* Prinsloo and Anil Kanjee (2005), HSRC**

The Human Sciences Research Council (HSRC) was contracted to develop the M&E framework of the programme, and to conduct the baseline (2000), mid-term (2002) and summative evaluation (2004) of the programme. At mid-

term the project evaluation framework was re-structured and only in 2002 control groups were integrated. This allowed for quasi-experimental (double difference method) to be applied to the summative evaluation, however this would capture only partially the impact of the intervention as the QLP had started already 2 years before. Control samples were chosen in the same districts and matched to programme samples based on similar social-economic and school variables. However this also meant that the control schools also received some level of partial treatment as support to district offices was also a component of the overall project. The implementation of the programme was also not homogenous as different service providers with their different approaches and quality of services were engaged in different districts.

Learners in both treatment and control groups in grade 9 and 11, were assessed in mathematics and English reading and writing through pre- and post- tests developed by the HSRC. Further analysis was done on the results from the 12 grade NSC exams, however this analysis was done at school-level rather than learner level, resulting in the sample sizes being too small to yield statistically significant results. The analysis of NSC results was also made more problematic, as the authors reported that programme schools were starting from a previously low base, and under scrutiny the schools could have also deliberately held back learners from entering grade 12 or taking demanding subjects in order to keep school NSC pass-rates high.

As part of the study the authors also conducted some impressive path analysis exercise to observe the effects of interactions of different home, school, class-room and contextual factors on the learning results of participants of the QLP. The study concluded that the QLP was overall successful and had a significant impact on the learning outcomes of the participating high schools. For the limitations described above however the effects are probably under-estimated, and the results from the analysis of the NSC pass rates are probably not reliable enough to be included in this meta-analysis. *Risk of Bias: 3*

Intervention 12: Plus Time

Plus Time was a learner-centered initiative funded by the Shuttleworth Foundation. It aimed to improve mathematics and English abilities among grade 8 learners in the Western Cape. It involved providing participants 20 hours of tutoring classes after school in the two subjects. It was piloted first in 2007 and then repeated again in 2008.

Evaluation 12: Cas Prinsloo (2009), HSRC

The HSRC and the Western Cape Department of Education collaborated together to gather learner test scores through the Continuous Assessments (CASS) and the Common Task Assessments (CTA). Learners were well matched between programme and control groups however sample sizes were relatively small. Results from 2007 assessments were inconsistent and revealed that many background and contextual factors influenced results. The results from the 2008 follow-up were much more promising with big gains made by participants in mathematics. Learners who were second language English-speakers benefited more from the programme than those who were mother-tongue English speakers. The author was very collaborative in providing missing statistics and further clarifications to the reviewers in order to include the study into the meta-analysis. *Risk of Bias: 3*

Intervention 13: Family Maths School Project

The Family Maths Project was an initiative designed by COUNT and funded by the Shuttleworth Foundation. It aimed to improve mathematics learning outcomes in primary school. Its underlying philosophy was to engage actors in the classroom as well as at home to maximise learning results of young children. Essentially the project provided educators training on innovative approaches to teach mathematics in early grades. It also facilitated a number of workshops for parents and caregivers, to support better the learners at home in their understanding of mathematics. The programme was piloted in Gauteng, Limpopo and Mpumalanga during the course of 2006.

Evaluation 13: SAIDE (2007)

The South African Institute for Distance Education (SAIDE) was commissioned to undertake the impact evaluation of the COUNT family project. The authors of the study demonstrated a strong grasp of the complexity of conducting quasi-experimental evaluation for causal inference of programme impact. Programme and control groups were carefully chosen and matched on similar geographic, social-economic and school factors, but also care was taken to not select schools close to each other in order to avoid spill over from treatment to control groups.

Although well designed, the implementation of the quasi-experiment suffered a lot of challenges. The mathematics test scores used by the evaluators were the standard school assessments rated by the teachers themselves, which could have caused some subjectivity and bias. Furthermore the study was unexpectedly challenged by the fact that many schools became hesitant or even refused to provide the learner results to the researchers. There were also other major data constraints, such as a large number of 'no-show' or attrition of learners between pre- and post-test, and in some cases principals deliberately switching the participating learners between the two tests. These testing biases were also coupled by programme implementation challenges. In some cases parents workshops never occurred. In other cases the wrong set of educators (not the ones teaching the treatment learners) took part in the teacher-training.

Having acknowledged all the above challenges and limitations of the experiment, the authors decided to exclude from the analysis all the 'problematic' cases in both programme and control groups, and followed only the individual learners which they had results for both pre- and post-test and no major problems occurred during implementation. Although this increased the accuracy of the results, it also narrowed the impact to the ATT. With the exclusion of many cases, less localised analysis could not be done anymore. The sample sizes had to also be significantly reduced, thus increasing the level of standard error. Nevertheless, with all the caveats and limitations, the evaluation was still able to prove with a strong degree of confidence the effectiveness of the project, especially on grade 2 mathematics learning outcomes. *Risk of Bias: 3*

Intervention 14: Maths Centre for Professional Teachers (MCPT)

The MCPT was a two-year teacher-upgrading project implemented in the Eastern Cape (EC) and KwaZulu-Natal (KZN) funded by the Zenex Foundation and implemented by Maths Centre, with the objective to raise numeracy and literacy levels at the foundation phase of schooling.

Evaluation 14: Research Institute for Education Planning (RIEP), (2008), Zenex Foundation

The project was evaluated by the Research Institute for Education Planning of the University of Free State. Language assessments in KZN were conducted in isiZulu and isiXhosa in the EC. The study reported that learners in KZN performed better than the learners in the EC. This evaluation presented some serious problems and very high risk of bias. The sample of learners involved in pre- and post-test were completely different and there were no signs of efforts to match learners across control and programme group either. Sample sizes, especially at the pre-test were also very small, and lot of important information was missing from the report. *Risk of Bias: 5*

Intervention 15: Mother Tongue Literacy Programme

The Mother Tongue Literacy Programme was a programme funded by the Zenex Foundation and piloted by the READ Education Trust in KwaZulu-Natal (KZN) and the Eastern Cape. Through the experience and learning emerging from the Learning for Living Programme (see above), READ highlighted that the grade 1 literacy materials, currently taught in English, were inadequate for the linguistic situation of most South African primary schools, especially in rural areas. READ therefore proposed the production of new grade 1 literacy books in isiZulu and isiXhosa to facilitate language learning in earlier grades. READ undertook to also train the Grade 1 teachers on how to use the new mother-tongue material.

Evaluation 15: JET Education Services (2006), JET

The Joint Education Trust (JET) was commissioned to evaluate the pilot programme, including assessing the impact on learning results. The evaluation was limited to two districts in the Eastern Cape, where 24 schools had received the intervention. The evaluation used two different control groups of similar schools. One control consisted of four schools who had not received any intervention, and the other control group consisted of four other schools who were receiving the IEP (other programme funded by USAID). JET-developed isiXhosa language assessment was administered in both project and control schools.

Generally speaking the counterfactual in this study is very weak as there is no clear evidence on a rigorous matching process which was undertaken in this quasi-experiment. The control group which the reviewer felt most appropriate to use was the group of schools and learners that were not exposed to other interventions (to avoid contamination). However, the sample size of this group was very small (N=45) to reach statistically significant results. Furthermore, contrary to the other studies in the meta-analysis, this evaluation reported the results of only one learning assessment during the project time frame. This did not allow for the classical DID model, which would allow for a reduction of the discrepancies between the 2 groups, and thus assist in achieving more accurate results. The risk of bias in this study is therefore fairly high. *Risk of Bias: 4*

Intervention 16: RUMEP Fort Beaufort Mathematics Project

In order to address the poor quality teaching in secondary school in the Eastern Cape, due to under qualified teachers with poor content knowledge and weak pedagogical skills, the Zenex Foundation launched a 3 year programme called RUMEP Project. The logic of the project was to improve learner achievement through improvement of the quality of teaching. This was done through enrolling mathematics high schools educators in a

B.Ed programme, and supporting them with additional resources and materials, classroom mentoring and workshops for peer exchange.

Evaluation 16: Paul Hobden & Sally Hobden (2009), QPiE

Quality Programs in Education (QPiE) was commissioned to conduct the summative evaluation of the programme. The evaluation consisted of qualitative case studies as well as a quantitative quasi-experimental analysis, utilising a control group of learners from schools from similar locations, performance and social-economic context. However it was challenging to have a clear overview of the evolving qualifications and training which teachers in the control schools would receive during the project time-frame. The evaluators had prepared a series of testing instruments to measure grade 10 Mathematics basic skills, in both February (pre-test) and October (post-test), as well as a Mathematics content assessment administered at the end of each year. One of the challenges of the implementation was the very high teacher mobility both within and throughout the different schools in the region. In order not to jeopardise the experiment the evaluators had to reduce the samples to schools in which less changes occurred throughout the project.

The evaluation found that although RUMEP professional development programme was implemented to high standards, the project had very little impact on learner achievement. The authors speculated to various problems in the project logic, as well as teacher and school factors having an influence on poor outcomes.

For the purpose of the meta-analysis, the authors were very cooperative in providing original data and clarifications from their original study. The reviewers utilised the effects sizes from the grade-10 basic skills tests of both 2007 and 2008 (which showed different results) as well as the time-series comparison of the mathematics content test between 2006 and 2008 (which involved different sets of learners). The re-calculations done in the meta-analysis found that the results were not as grim as reported in the original study, however there were indeed some serious problems with both implementation and the counterfactual (i.e. potential selection bias), which would thus not provide conclusive evidence on causality and if the intervention actually did have any impact on the improvement of learner results of participating schools. *Risk of Bias: 4*

Intervention 17: Mindset teacher development programme

Between 2006 and 2007, the Zenex Foundation funded the Mindset Network to run a programme in collaboration with the North-West Department of Education.

The Mindset programmes had as its objective the improvement of teacher capacity in secondary schools, especially in the subjects of mathematics and physical sciences. The programme would provide multi-media equipment, Mindset mathematics and science video and print material, and in-service teacher training. The logic of the programme was that educators would integrate Mindset resources to enhance their teaching and thus contribute to better learning results in mathematics and science at the grade-12 NSC exam.

Evaluation 17: JET Education Services (2007), JET

The Joint Education Trust was commissioned to conduct the evaluation of the programme, which included an assessment of the quality of the materials produced, change in classroom teaching practices and the impact on learner performance. This involved both qualitative case studies as well as the design of a quasi-experiment. From

the two districts in the North-West involved in the project, 18 schools were selected as project schools and six other random schools were selected to be the control group. Subsequently 25 learners were selected from each grade in both treatment and control schools to participate in the experiment.

JET designed the mathematics and science assessments that were used in June 2006 (Baseline), February 2007 and August 2007. The evaluation reported a cyclical trend of improvement in all grades and subjects that were assessed. Serious concerns however were raised with the overall low results in both programme and control groups, where majority of learners didn't meet minimum marks to pass to the next grade.

In order to calculate the effect sizes for the meta-analysis, the reviewer utilised the data contained in the report to make comparisons between the grade 10 and grade 11 mathematics and science results between mid-2006 and mid-2007 (averages were made between February and the August 2007 results). The reviewer also undertook some cohort analysis where the same learners were followed in their progression from grade 10 to 11. With both these methods of analysis, there appeared to be a large amount of attrition between participants of pre- and post-tests, thus posing some high risk of participants bias.

Aside from the fairly weak approach to the counterfactual, the different project schools recorded very heterogeneous degrees of usage of the equipment (as registered by the logon registries). The project also suffered from major implementation challenges which jeopardised the experiment. During the time-frame of the project, three schools suffered from theft of equipment, there was a major national teachers' strike during that year, followed by the implementation of school recovery plans, which took most of the attention of teachers and principals. There was also a mid-project change in the INSET service provider. Overall the quasi-experiment had too many challenges and a high risk of bias, which wouldn't allow us to utilise the results with confidence. *Risk of Bias: 4*

Intervention 18: Matchs Centre/NMMU teacher training

Between 2005 and 2007 the Zenex Foundation funded two teacher-training programmes. One implemented by ELET and UKZN addressing literacy in 30 primary schools in the Eastern Cape and KwaZulu-Natal. The other was a mathematics programme implemented by Maths Centre and the Nelson Mandela Metropolitan University (NMMU). The three-year mathematics programme was targeted at teachers in 19 secondary schools around Polokwane, in the Limpopo province. The teacher development programme consisted in a B.Ed qualifications upgrading at NMMU, coupled by supplementary INSET workshops, provision of teaching resources, coaching and mentoring of educators by the Maths Centre.

Evaluation 18: Feedback Research & Analytics (2008), Zenex Foundation

Feedback Research & Analytics was commissioned to conduct an evaluation of only the mathematics component of the Zenex programme, which looked at impact on teacher knowledge and practice, but also on learner results of the participating schools in Limpopo. The evaluators development a mathematics assessment instrument which was delivered to grade 10 learners in March (pre-test) and in October (post-test). 10 additional classes of grade-10 learners in neighbouring schools were also administered the test in order to provide a control sample. There was generally a large amount of attrition between pre- and post-test (between 24% and 38%), therefore only matched results for learners who took both the March and October tests were used. The same quasi-experiment was

undertaken in both 2006 and 2007 years. The evaluators also examined the various school factors which contributed or not to learner achievement

The results of the 2006 experiment, did show a small but statistically significant impact of the programme. The 2007 repeat, however, showed the contrary, where the control group outperformed the programme school. Overall the results of the grade-10 mathematics assessments across all the groups was very poor and showed that only 20% of learners had achieved the basic competencies to pass to the next grade. The study demonstrated that the support provided by Maths Centre/NMMU/Zenex did not seem to make a big difference in the general provincial trends.

From a methodological perspective, the counterfactual of the evaluation was not very meticulously designed and opened the possibility for selection bias. There were certainly some outliers, such as the high performance of the Harry Oppenheimer Secondary and the sudden decline of Mmamolope, Mmankogaedupe, Ncheleng, and Pholeka schools in the programme group, which might have skewed the results of the study. *Risk of Bias: 3*

Intervention 19: Platinum mining contribution to education in South Africa

The below study was done on the platinum industry, the largest export of South Africa, and its contribution to human capital and social development in the country. The study actually looked at three different but interconnected interventions. They will thus be described in three different levels:

Intervention 19.1: Presence of the mine

Large reserves of platinum ore are present in different areas of the Limpopo and North-West province of South Africa. Whenever mining operations start in an area, this generates employment and stimulates local economic activities, however it also causes disruption, and introduces environmental hazards and other social problems for the local communities.

Intervention 19.2: Anglo Platinum Education Programme

As part of its Community Engagement and Development (CED) strategy and its Social Labor Plan (SLP), Anglo American Platinum (the biggest corporate social spender in South Africa) invested between 2009 and 2013 around R100 million in education and skills development programming for the communities around their mining operations in Limpopo and North West. These consisted of a wide range of interventions from early-childhood development (ECD) to adult education, from infrastructure, LTSMs to trainings, interventions aimed at teachers and learners, and support offered to school management and districts offices.

Intervention 19.3: Radical Maths & Science Winter Camps

One of the numerous interventions implemented by Anglo Platinum was a two-week intensive workshop during the school holidays to help grade-12 learners to prepare for the NSC exam, and improve their results particularly in mathematics and physical science, critical subjects needed to equip the potential future technicians and engineers that the company would recruit. In both 2011 and 2012, hundreds of learners were gathered from the schools around the different mining operation areas, and brought together in an intensive winter camp during August/September break, facilitated by expert teachers from Radical Maths & Science organisation.

Evaluation 19: Neissan Besharati (2014), SAIIA (see also previous chapter 4)

In order to assess the impact of all three of the above interventions, Neissan Besharati, utilised a mix of PSM (Jalan & Ravallion, 2003, Dearden et al., 2008, Rosenbaum & Rubin, 1983) and DID method. The analysis was done using a large dataset received from the Department of Education of over 1800 schools. Treatment schools were carefully matched based on previous performance, school resources and other social-economic demographic variables to the control schools. Learning results were measured through time-series data from the schools' NSC results, with 2008 acting as the pre-test and 2012 as the post-test. To emulate the standard grade (SG) and HG of the pre-2007 NSC dispensation, the authors created an indicator for normal passes (above 30% score) and an indicator for University quality passes (above 50% score) for both mathematics and physics subjects, similar to the normal pass and Bachelor pass rates used for the overall NSC results. As part of the study, qualitative field-work was also undertaken to understand the dynamics of the interventions and the local population, as well as econometric analysis (with over 100 variables) to observe which factors were most correlated to improved mathematics and science results.

Besharati discovered that schools situated close to the mining areas had a systematic decline in learning results, compared to the schools which were farther away from the mines. Impact assessment was subsequently done on Anglo Platinum overall FET programme to see if the company investments were able to reverse such trends, however the results showed no statistically significant improvements in the learning outcomes of the schools participating in the Amplats programmes compared to the control schools. A more focused analysis was subsequently conducted on the impact of the specific intervention of the Radical Maths & Science Winter Camps on the 2011 to 2012 NSC results. Although these yielded positive effects at the learner level, when taken to the overall school scores at NSC, the results were inconclusive and had too high standard errors due to small sample sizes. The study concluded that although few selected beneficiaries appreciated the Anglo Platinum programme, the interventions failed to translate into significant gains for the general school system in Limpopo and North West province. The very small effect of Anglo Platinum education interventions were out-weighted by the large negative effects that the mere presence of the mine caused on learning outcomes of the surrounding schools.

For the purpose of the meta-analysis only the effect sizes for the mathematics, science and overall pass rates for each of the three interventions were extracted and calculated from this study. As the author of this evaluation is also one of the authors of this meta-analysis, this specific study was reviewed by other external assessors. *Risk of Bias: 2*

Intervention 20: Education Quality Improvement Partnership (EQUIP)

EQUIP is a public-private partnership run by the National Business Initiative (NBI) aimed at improving the quality of education in disadvantaged and underperforming schools of South Africa. The programme started in 1995 with a small pilot in KwaZulu-Natal and after a decade it gradually expanded to more than 500 schools in 8 provinces of the country. The programme was supported over the years by more than 50 private companies. At the heart of its model, EQUIP provides an enabling framework to promote whole school development through local leadership and empowerment. A key component of the programme is the development of school development plans (SDPs), which guide school improvement and can be used to mobilise support and partnerships from broader stakeholders. At

different stages of its evolution, the EQUIP programme included school management and governance strengthening, in-service teacher training, up to full whole school development. EQUIP has also provided direct support to the national and provincial departments of education by supporting district development and contributing to other policy processes such as the Integrated Quality Management System (IQMS), Norms and Standards for School Funding and the National Curriculum Statements.

Evaluation 20: Eric Schollar (2002), ESA

Throughout the decade there have been numerous evaluations and reviews of the EQUIP programme (Mouton, 2004, Schollar, 2004, Hampel et al., 2006), however this meta-analysis was able to include only Schollar's 2002 evaluation of the EQUIP pilot project (1997-2001), which was implemented in 20 schools in Gauteng, KwaZulu-Natal and Western Cape, and involved only the basic intervention of SDP, strengthening of school management and some district development. Schollar's evaluation involved a longitudinal quasi-experiment which sampled groups of learners from control and treatment schools matched on the usual demographics, school characteristics and social-economic similarities. ESA testing instrument were used for the baseline (1998), mid-point and post-test (2001) to test literacy and numeracy in grades 5 and 7, as well as to follow cohorts of learners (from grades 4 to 6, and from grades 5 to 7). The quantitative findings indicated that EQUIP had a significant impact on learner achievement. The study was also complemented by qualitative analysis, school visits and interviews which confirmed that EQUIP had contributed to a change of management culture and efficiency, and recommended the programme to integrate INSET teacher professional development as part of the next phases.

As usual Schollar was very forthcoming in providing the original dataset of the study from which to extract the standard deviations, means and sample sizes required by the meta-analysis. The reviewers noticed some sizeable attrition occurring in the cohort analysis and also a small reporting error in the results of the grade 7 post-test, which was corrected by referring back to the original dataset. *Risk of Bias: 3*

Intervention 21: COUNT Numeracy programme in Uitenhage

Between 2003 and 2005, the Zenex Foundation funded COUNT to implement a numeracy programme in 20 schools of Uitenage town in the Eastern Cape. The programme targeted the foundation phase learners (grades 1 to 4) and consisted primarily in a teacher training programme coupled with some in-school mentoring and provision of teacher support material. The COUNT Numeracy programme aimed at improving learner numeracy results through the improvement of educator knowledge, pedagogical practices and attitudes.

Evaluation 21: Roberts & JET (2006)

The final evaluation of the COUNT Uitenhage project was conducted by Jennifer Roberts and JET Education Services in 2005/2006 using a mixture of qualitative methods (including 10 project schools) and quantitative methods (involving 6 project and 2 control schools). JET designed mathematics assessments that were used to test cohorts of grade 3 learners at the beginning and at the beginning of the year, as well as to compare the results of grade 3 treatment learners at the end of 2004 with grade 4 non-treated learners at the beginning of the same year. The results from the qualitative assessment showed that the teachers appreciated the teacher development programme offered by COUNT, however these did not translate into improved learning outcomes of the learners

affected by the programme. There was in fact no statistically significant difference between the learner results in the treatment and in the control group.

The quantitative evaluation, tried to emulate a pre- and post- quasi-experiment, however the matching criteria for the control and programme schools was unclear, presenting a very weak counterfactual. What was apparent is that the control group contained an outlier school, that performed exceptionally well compared to the other schools in the experiment. Sample size was also very small and did not provide enough power to offer reliable results. Although the programme started in 2003, the baseline and the post-test were both conducted in 2005 thus the baseline was 'contaminated' by the fact that the teachers had already been exposed to 2 years of intervention at the time of the pre-test. There was also some learner attrition between the baseline test and the follow-up. Also considering the short time span between the pre-test and the post-test in the final year, little impact can be expected to be detected from such flawed experiment. To some extent the study can also be criticised for poor external validity as it focused only on schools in Uitenhage (close to the Volkswagen factory), which presents a very specific social-economic reality than other small towns in the rest of the country. *Risk of Bias: 5*

Intervention 22: Public School Maths Challenge Programme

The Match Challenge Programme of the Epoch & Optima Trust was a unique programme that targeted the top-end public schools throughout the country. Contrary to the approach of most other donors, Epoch & Optima Trust focused on the high-performing schools with strong track record of delivery, where interventions could maximise impact. Top-performing high schools were offered conditional grants between 0,6 and 1,4 million rand to spend in what the schools deemed most appropriate to improve both the quantity and quality of mathematics matric passes, particularly among the previously disadvantaged populations. The majority of the grants were usually spent in increasing and strengthening the school's human resources dedicated to mathematics - additional educators, teaching assistants and teaching hours. Other school expenditures would include the procuring of LTSMs and technological equipment (computers, workbooks, interactive white boards, ipads, etc.) or on additional learning activities (revision and remedial classes during school breaks, weekends and after school). Between 2008 and 2012 over R 68 million was disbursed to 64 schools benefitting in total 58,747 learners.

Evaluation 22: Schollar & Mouton (2014), ESA & ERA

The Maths Challenge programme was evaluated by two of the most seasoned education evaluators in South Africa. The study consisted of a mixture of qualitative work through field visits, observations and interviews with school principals and teachers, as well as quantitative analysis along geographic, social-economic, racial and provincial parameters.

A quasi-experimental design was also used to observe improvement in learner performance, by utilising NSC exam results provided by the DBE, including proportion of learners writing mathematics, the pass rates at 30, 50 and 60 percent, as well as average weighted scores in mathematics. Analysis was conducted at the school level rather than the learner level. All 64 programme schools were matched to control schools from the DBE database using parameters such as school size, social-economic demographics and past performance and enrolment.

The study showed that there was a statistically significant improvement in learner results of programme schools compared to outcomes of control schools. The grant provided by Epoch & Optima contributed to 812 more learners (including African and previously disadvantaged learners) to pass the NSC exam at a level sufficient to enroll into University. The qualitative findings highlighted that this was thanks to the increase of human resources and teaching time which the awarded schools could afford as a result of the grant. With a per-capita expenditure of R1158 per learner, the evaluation concluded that the project provided good returns on investment for the two foundations.

Several critics of this evaluation have highlighted that there was a strong selection and motivation bias underlying this study. Schools that took part in the programme were selected based on a mixture of criteria which was not always clear and consistent, but the ones that benefitted from the grants were clearly the stronger and more enthusiastic applicants which would have performed better anyway compared to other schools in their category. The counterfactual was thus fairly weak. One could argue also that the external validity of the programme results is also very limited as the intervention was done with the 'crème de la crème', which does not reflect the generality of the schools of South Africa. Finally, there was a vast heterogeneity in the treatment, as the the grant was of different amounts and utilised in different ways by different schools. But otherwise the analysis was very thorough and the data presented in the report clear and comprehensive. *Risk of Bias: 3*

c) Academic experiments

A number of different scholars from different South African universities have over the years conducted experiments to test different educational theories, hypothesis and models. In such experiments the interventions were driven by the researchers, and in many cases the treatment consisted of specific technology, LTSMs or particular approaches to teaching and learning.

Intervention 23: Numeric / Khan Academic after-school programme

The Potter Foundation and RUBEN sponsored a RCT conducted by the University of Cape Town in collaboration with the Western Cape Department of Education and Innovations for Poverty Action (IPA). The experiment involved introducing an after-school computer-aided numeracy programme to 11 classes of grade 8 learners in 9 schools of Cape Town Metro area. The programme was implemented by Numeric and consisted of the use of mathematics coaches to facilitate twice a week after-school mathematics tutorials, using didactical videos and online computer games (knowledge maps) produced by the Khan Academy. The pilot programme ran from January to November 2013, for the interested and dedicated learners from grade 8 of the selected Schools in Cape Town metro area, which met the requirements of adequate computer labs and internet connectivity.

Evaluation 23: Bohmer (2014), UCT

The results of the RCT above were drawn out from the masters thesis of Bianca Bohmer, which was part of the UCT/IPA evaluation team. The experiment was very carefully conducted by randomly assigning half of the 472 applicants and eligible candidates into the treatment group and the other half into a control group, stratified in the 9

schools and by baseline mathematics scores, creating a very powerful counterfactual. The experiment was also very carefully and tightly administered, ensuring high participant attendance, compliance to the treatment, and reducing to a minimum spillover effects and attrition. Both control and treatment learners where tested at baseline, mid-point and end-point on a numeracy test and a mathematics grade 8 test, developed by an independent assessment committee, consisting of an IPA consultant and 2 high-performing Western Cape teachers.

The RCT proved that the Numeric after-school computer-aided programme (using Khan Academy ICT material) had a significant impact on the numeracy results of the treated learners. Further econometric analysis was performed also to test heterogeneity of the treatment effect through gender, race, home language, school quintile and other variables. The only factor which had a strong correlation on mathematics achievement was the learner English ability, which confirms the findings of many other studies in South Africa.

The main criticism of this experiment is that the results have very low external validity as the experiment was conducted with the most eager, committed and motivated learners/families, from a group of selected well-resourced, ICT-equipped, and well-functioning schools in Cape Town metro area, which is far from representational of the broader reality of the South African schooling system. *Risk of bias: 2*

Intervention 24: Problem-based Learning (PBL)

Problem-based learning (PBL) is a teaching strategy that is learner-centred, hands-on, activity- and inquiry-based and makes learners more accountable for their own learning (Savoie & Huges, 1994, Vernon & Blake, 1993, Hattingh & Killen, 2003). Attempts to use such learner-centered strategies, especially PBL, have often also attracted criticism that learners acquire less content information than in direct lecture-based strategies (Gallagher & Stepien, 1996). This introduces the age-old debate of depth versus breadth of curriculum content coverage.

Evaluation 24: Annemarie Hatting (2003), SAJE

Annemarie Hatting of the University of Pretoria challenged the above assumption by conducting an experiment among grade 10 science learners in three average-performing public high schools in Gauteng and one in Mpumalanga. Each school principal supplied 2 grade-10 classes for the experiment. One class would be taught using PBL and the other would be taught through traditional 'chalk and talk' approach. The researcher developed and moderated a tailor-made science test to measure learner scores at both pre- and post-tests. To create statistically similar and comparable groups, learners were sampled for the treatment group and matched to others in the control group based on same gender, school and test-scores during pre-test. Of the initial 202 learners in the four schools, only 70 experimental and 70 control learners complied with these criteria and featured in post-test comparisons.

Science classes occurred on average three hours per week. The same teacher was responsible for teaching both the PBL and the control group in a particular school. The teachers in the four schools all had exactly the same qualifications, had received six months pedagogy training by the researcher, and had the same instructional materials and resources, for the two different interventions they had to implement. At the end of the experiment, mean scores of learners being taught with PBL were slightly higher than those taught through traditional lecture-based methods, however the difference was not statistically significant. The results for the least rejected the common assumption that subject content is sacrificed when PBL methods are used in teaching. Further qualitative

analysis was conducted in the study to probe the levels of enjoyment by both teachers and learners who were exposed to PBL.

Statistics were well-reported in the paper and could be extracted easily for the meta-analytical calculations. As the sample sizes were unfortunately too small, results have high standard errors, therefore no strong conclusions can be deduced. The author of the study was very cooperative in providing further clarifications and information to the reviewers. *Risk of Bias: 3*

Intervention 25: Mathematics Incubator School Project (ISP)

The Mathematics Incubator School Project (ISP) was initiated in 2004 to address the shortage of qualified mathematics teachers in secondary schools in the Eastern Cape. By 2007 it evolved into the production of a series of DVDs which would cover different aspects of the mathematics syllabus. The DVDs would be used as a powerful teaching and learning aid, which was given to teachers and participating learners. The DVD would be used through a blended approach involving, additional Saturday sessions, facilitator discussion, peer interaction and tailored assessments.

Evaluation 25: Padayachee, P, Boshoff, H, Olivier, W & Harding, A (2011), Pythagoras

The authors of the study conducted a quasi-experiment on the blended DVD approach of the ISP, by conducting a case study in one school where pre- and post- mathematics scores of 6 ISP participating learners were compared to 14 control learners in the same school. The results of the experiment showed that the ISP had a very large impact on the mathematics results of participating learners.

The reviewer found this study to be extremely problematic. Firstly, there was clearly indication of a selection bias. Authors made no attempt to create a credible counterfactual. They themselves reported that participants of the programme were selected among the best-performing learners who had interest in pursuing science, engineering and mathematics subjects at University, thus were clearly different learners from the rest of the learners in the control group. Furthermore the sample sizes used were way too small to conclude any statistically significant results on the impact of the intervention. *Risk of Bias: 5*

Intervention 26: Language companion

Extensive literature (Adler, 1995, Lemke, 1990, McLean, 2000) explain how in multi-lingual societies, such as South Africa, the challenges in learning mathematics often come from simple language barriers. A group of researchers from Tshwane Univeristy of Technology (TUT) developed an electronic learning and teaching tool (the 'learner companion'), which would translate English mathematical terms contained in the grade 1-4 syllabus into isiZulu, isiXhosa, Setswana, Sesotho and Afrikaans, as well as providing a graphical representation of the word. The aim of the learner companion was to assist in overcoming language challenges and encourage better understanding of mathematical concepts, which are usually explained only in English.

Evaluation 26: Hendrick Botes and Andile Mji (2010), SAJE

The researchers set up a quasi-experiment to test the efficacy of the learner companion in a sample of schools in the North-West province. 10 treatment schools and 10 control schools were carefully selected and matched based on

demographic and social-economic parameters. A total of 1,164 leaners were given the language companion and their educators trained and encouraged to use the tool as part of their teaching. Another sample of 1,184 learners together with their respective teachers in the control group would be observed while they continue business as usual by running mathematics classes in English without the specific language aid. Care was taken by the researchers to implement the intervention as faithfully as possible and avoid any form of 'contamination' between the two groups. The authors developed their own mathematics test instruments that were administered by the teachers in both treatment and control groups in February (pre-test) and November (post-test).

Through a well-executed experiment, the study demonstrated that the 'learner companion' had a statistically significant effect on the mathematics scores of treated participants. Further qualitative analysis was conducted by the authors through interviews with the teachers so to confirm the quantitative findings, as well as to receive suggestions for the improvement of the electronic language aid. The results were very positive, however caution needs to be applied when considering the external validity of the study and a potential Hawthorne effect as both control and programme teachers were aware and closely collaborating in the implementation of this experiment. *Risk of Bias: 3*

Intervention 27: 'Back to the Basics' Workbooks / PMRP

As evidenced above, one of the most prolific evaluators in South Africa's education sector has been Eric Schollar. Schollar has been engaging with quasi-experimental evaluations in education for nearly two decades, in a sector which, especially in the beginning, was highly dominated by qualitative researchers.

Based on the evidence and his accumulated experience in the evaluation of many education programmes for different parties, Schollar decided to develop a programme that would address the causes of the very poor mathematics outcomes in South African public schools. As such, with support of National and Provincial Departments of Education the Primary Maths Research Programme (PMRP) was initiated in 2004. The programme was preceded by extensive empirical research in primary schools in Limpopo that would provide the evidence-base for the next stages of curriculum development.

ESA developed a new set of learner workbooks called 'Back to the Basics', which would assist learners to move away from 'unit counting' and to conducting with ease the basic arithmetic operations prescribed by the curriculum. To assist in multi-grade classes (which are very common in South Africa), the workbooks would begin with a diagnostic test which would place learners at the appropriate level (grade 3 to 6). The workbooks would contain different lessons, which would elaborate on mathematical concepts, applications, repeated exercises and regular self-assessments.

In 2007 the Limpopo Department of Education agreed to test the materials in 3 circuits of the Vhembe district. What followed was a 1,5 day training for all principals and teachers involved in the experiment. Mathematics teachers were provided a teachers manual, prescribed daily lesson plans, test instrument, an implementation diary, and copies of the Back to the Basic workbooks for all their learners. The materials covered a 14-week programme of 70 lessons for grades 3, 4, 5 and 6. Researchers conducted visits to each school at least twice to check on proper implementation of the programme.

After very positive findings emerging from the first pilot (see evaluation below), the next phase was to transfer the operational management of the PMRP fully to the Limpopo Department of Education, who would take the project to scale by distributing the workbooks to all the other 125 primary schools (20,000 learners) of the Malamulele cluster between 2010 and 2012.

The PMRP, was a multi-phased evolving action research project which was funded at different stages by the Zenex Foundation, Shuttleworth Foundation, Anglo American Chairman's Fund and Xstrata South Africa, implemented by ESA in close collaboration with the Limpopo Department of Education, with the purpose of incorporating learning into education policy in South Africa.

Evaluation 27.1: Eric Schollar (2015), UCT

The Primary Maths Research Programme, was a multi-phased action research effort led by Eric Schollar, that included 5 evaluations, including a large RCT. Several evaluation reports and studies exist of this programme, but the main documents that will be referred to in this meta-analysis are Eric Schollar's final report of the 2007 randomised experiment (Schollar, 2008) as well as his PhD (Schollar, 2015) which brings together the decade-long research endeavours in Limpopo province.

The first RCT on the Back to the Basics workbooks was conducted in 2007 in collaboration with the Limpopo Department of Education. 20 schools from the Vhembe district were randomly selected to partake in the intervention, and 20 others were randomly selected to be part of the control group. The researchers were very closely involved with the schools and made sure that the treatment was implemented correctly. ESA pre- and post-tests were administered in both grade 4 and grade 6 to assess the impact of the workbooks. Separate results were calculated for learners who benefitted from different 'dosages' of the programme – at least 7 weeks (50% of the intervention) and 11 weeks or more (80% exposure). The findings of the experiment were that the workbooks had a phenomenal impact (up to 130%) on learner performance, especially the ones which had utilised the material for at least 11 weeks. The gain in scores achieved by the programme were in the order of 2 or 3 times the kinds of learning gains effected by other school intervention experienced in South Africa (Taylor, 2007).

In 2010 before the programme was extended to the rest of the district, another follow-up experiment was undertaken with the same cohort of participants to see if the effects of the initial intervention would continue to persist three years later. The initial grade 4 learners in control and programme groups in 2007 (particularly the ones who had received high-dosage treatment) were re-tested again in 2010 while they were in grade 7. Notwithstanding a 3-year withdrawal of ESA from the participating schools, the programme group seemed to still yield significantly higher mathematics results than the control group.

The second experiment tried to address the potential presence of a Hawthorne effect, which was very likely to occur in the first experiment, due to the close proximity of the researchers to the participants. However, in the follow-up experiment, there could have also been some traces of selection bias, as the treatment group was constituted of the most motivated and best-performing participants who had already proven their commitment by participating in more than 11 weeks of the programme (measured as intention-to-treat effect). In upscaling to education policy setting however 100% exposure to the ESA treatment would be unlikely. In his research, the author explained the likely presence of spillover effects and contamination from other interventions in the district,

which could have compromised the implementation of the experiment. Eric Schollar's work on the PMRP has also been repeatedly criticised for high potential of evaluator bias, as the author was not only the lead evaluator but also the designer of the programme, therefore had a strong vested interest in proving its success. Schollar was very forthcoming in providing the original datasets of the 2007 PMRP experiment and 2010 follow-up. Although results for grade 4 and 6 were collapsed and reported jointly in the PhD, the reviewers could re-calculate the separate effect sizes for the impact on each of the grades, which were important for the sake of comparisons within the meta-analysis as well as with Fleisch et al. (2011) in the subsequent evaluation of the PMRP workbooks (see below). *Risk of Bias: 3*

Evaluation 28.1: Brahm Fleisch, Nick Taylor, Roelien Herholdt and Ingrid Sapire (2011)

Having witnessed the outstanding results of the Back to Basics workbooks in Limpopo, a group of researchers from Wits School of Education and JET Education services decided to replicate the experiment in Gauteng with funding from the DG Murray Trust. The team wanted to find out if providing the PMRP workbooks would improve learner mathematics achievement more than conventional textbooks.

In 2010, the authors conducted a RCT in Gauteng province, where 44 low-income schools were selected to participate in the study, half of whom would be randomly assigned to the treatment group while the other 22 to the control group. Pre- and post- test for grade 6 mathematics learners were conducted using an assessment instrument developed by the research team. The evaluators delivered the Back to the Basics workbooks to the treatment group following the same procedures as the original PMRP Limpopo experiment. For ethical reasons, rather than giving nothing to the control group, the authors decided to provide to the other schools a carefully chosen mathematics textbook (Classroom Mathematics), as 'enhanced standard practice'. As in the Limpopo study, assessments were also made on the impact of the intervention, for the sub-set of learners who covered over 79% of the material (high dosage).

The findings of the experiment showed that both programme and control group made significant progress from pre- to post-test, however there was not much difference between the final results of the learners who used the PMRP workbooks and the ones which used the conventional textbooks. In analysing the subset of learners who covered more than 79% of the materials the learners with Classroom Mathematics textbooks performed significantly better than the learners that used Back to the Basics Workbooks. The study concluded that no policy warrant could thus be issued on the relative effectiveness of the Back to the Basics workbooks.

Overall this was an excellently performed RCT, however having lacked enough qualitative work, it resulted in a typical 'black-box' evaluation (Chen & Rossi, 1987, Bickman, 2000, White, 2009, Mouton et al., 2013), where not enough understanding could emerge on the why and how of the results of the experiment. What is striking nevertheless is that the impact assessment of the exact same intervention performed first by Schollar (2008) and then by Fleisch et al. (2011) yielded diametrically opposite results. One could argue that Schollar, as the original developer of the PMRP workbooks, had a more committed approach to the implementation of the programme than the later more independent researchers; or that the social-economic and institutional context of urban Gauteng is obviously very different than rural Limpopo. All of these could certainly be influential factors on the results of the two experiments. However, the biggest distinguishing factor between the two evaluations was

actually the study design. While Schollar provided nothing to the control schools, Fleisch et al. (2011) effectively gave an alternative intervention to what was supposed to be a control group. The WITS/JET study in fact did not prove that the PMRP workbooks did not have an effect, but simply that a well-used conventional textbook, such as Classroom Mathematics, has even more effect on learning outcomes than tailored workbooks, such as the Back to the Basics material. *Risk of Bias: 3*

5.6.3 Problematic studies excluded from the meta-analysis

As discussed in earlier sections, when conducting a meta-analysis care needs to be given to appraising the quality of the studies to ensure that problematic evaluations with weak methodology or faulty implementation do not skew or mislead the overall results of the analysis. As such all studies were carefully reviewed and rated for their degree of bias. A further screening of the above studies was conducted, to exclude any evaluation that had high (4) or critical risk of bias (5). This left only the well-implemented RCTs and strong quasi-experimental studies, such as the ones using RDD, PSM or other techniques that took care of selection bias. Evaluations with credible results normally were conducted with large samples and therefore had low standard errors and small confidence intervals. The studies that were not able to meet the quality criteria for the next steps of the meta-analysis were the following (Table 29). Further justification of the problems that led to the exclusion are provided in the previous descriptive section.

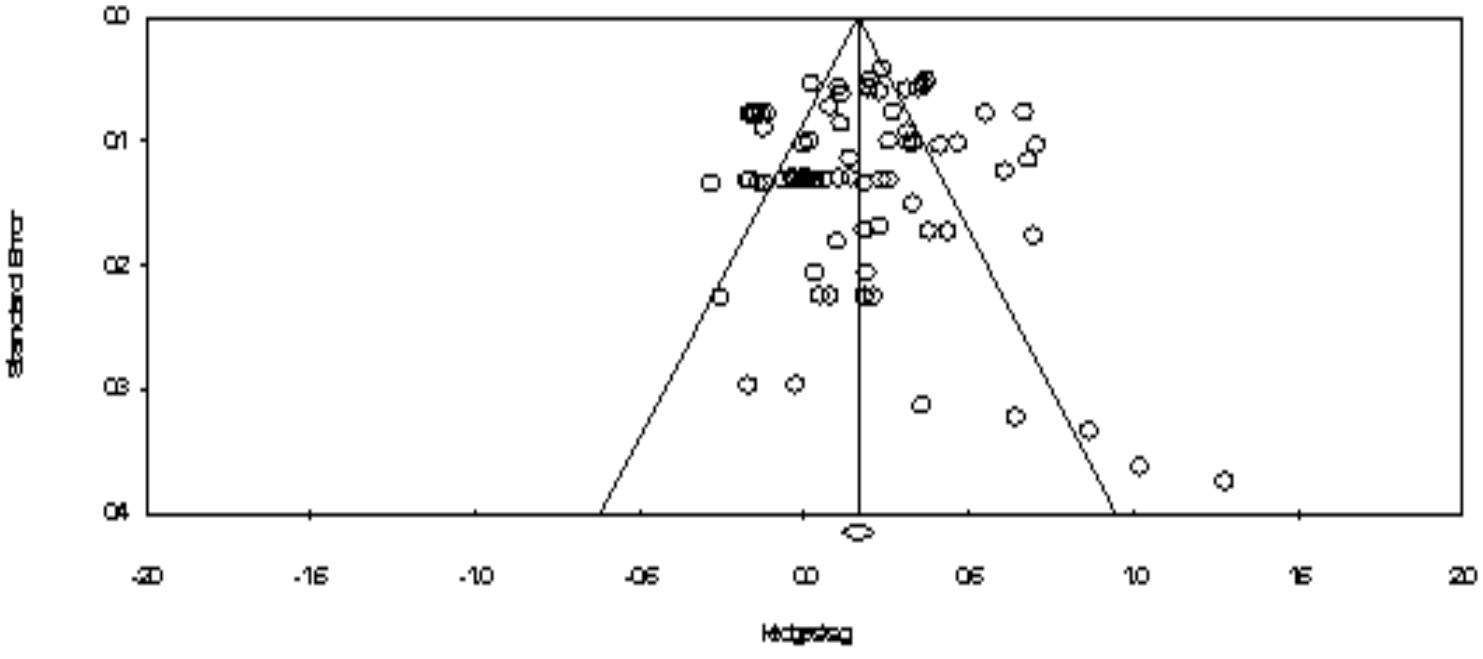
Table 29. *Studies Excluded from the Statistical Meta-analysis*

- JET Education Services (2006), Evaluation of the Zenex Funded READ Mother tongue literacy programme, Joint Education Trust
- JET Education Services (2007), Summative Impact Report: North-West Province, Zenex Foundation Educator Support Programme of the Mindset Network Project, Joint Education Trust
- De Chaisemartin, T (2010), CTLI Evaluation 2010 Report, Joint Education Trust
- Louw, Muller, Tredoux (2008), Time-on-Task, Technology and Mathematics Achievement, Evaluation and Programme Planning.
- Research Institute for Education Planning (RIEP, 2008), Report on the Impact Study Eastern Cape and KwaZulu-Natal on Literacy and Numeracy grade 4, University of the Free State.
- Hobden, P, Hobden, S (2009), RUMEP ZENEX B.Ed Project, Summative Evaluation Report, Quality Projects in Education

- Schollar, E (2001), A review of two evaluations of the application of the READ primary schools program in the Eastern Cape Province of South Africa, International Journal of Education Research, 35 (2001) 205–216
- Padayachee, P., Boshoff, H., Olivier, W., & Harding, A. (2011). A blended learning Grade 12 intervention using DVD technology to enhance the teaching and learning of mathematics. Pythagoras*, 32*(1), Art. #24, 8 pages.
- Gustafsson, M, Taylor, S (2012), Treating schools to a new administration the impact of South Africa's 2005 provincial boundary changes on school performance, Stellenbosch Economic Working Papers: 28/13
- Prinsloo, C, Kanjee, A (2005), Improving Learning in South African Schools, the Quality Learning Project Summative Evaluation Technical Report, Human Sciences Research Council ***(Only the effect sizes calculated from the NSC results were excluded; the results from the grade 9 and grade 11 assessments were still included in the meta-analysis)***

The screening process now left the meta-analysis with **18 credible studies of 19 different interventions**, which covered overall **82 different effect sizes**, as the main cases and units of analysis. This allowed for a legitimate comparison of results that had a certain level of reliability, and of studies and interventions that were placed on the same level plane, with similar contexts and evaluation parameters. In other words, 'apples and apples' could now be compared. The funnel plot (Figure 41) illustrates the effect sizes against the sampling errors in all the studies in the meta-analysis once the ones with high risk of bias had been removed.

Figure 41. Funnel plot of effect sizes by standard errors.



From the funnel plot (Figure 4) we can see that most studies are at the top, therefore the majority of the results reported are statistically significant, although some studies with a standard error below 0,1 provide slightly less reliable results than the evaluations done with larger sample sizes.

Funnel plots normally help in assessing the degree of publication bias, however we can also see that there is a similar number of South African studies that show positive as well as negative effects on both side of the zero effect line. The symmetrical nature of the funnel plot highlights that both studies reporting negative results as well as positive results were captured in this meta-analysis, and thus there is a fairly limited presence of publication bias.

There are however a number of major outliers at the right hand side of the plot, which indicate extremely and unusually high effect sizes. Most of these cases emerge from Schollar's evaluations (2015, 2014, 2005, 2002), therefore results from these studies need to be assessed with more caution. These anomalies warrant a closer look at the results and process undertaken in these particular studies, and a potential sensitivity analysis (Walker, Hernandez, & Kattan, 2008) when including them in the broader meta-analysis.

An eye-ball look at the above funnel plot, highlights also the sad reality that the majority of the education interventions implemented in South Africa had very small effects and at times even a negative impact on observed learning outcomes.

5.6.4. Adjusting standard errors to cluster sampling

As previously discussed, sampling has a big impact on the standard error and confidence intervals of the effect sizes reported by the various impact studies. Many, if not all of the RCTs reviewed in this chapter have used a statistical sampling technique referred to as clustered sampling when conducting the randomisation process. Clustered sampling is possible when the population can be divided into relatively homogeneous and balanced natural groupings (schools, cities, provinces, age or gender groups, etc.) (Trochim, 2006).

In the above studies and in other experiments conducted in the education field, participants are clustered in schools and classes. For practical, economic and ethical reasons, randomisation is often done at school or class level rather than learner level. This however compromises the sampling power and therefore the precision of the estimated effect size. It also affects the internal validity and the external validity of the study as findings drawn from studies done on fewer schools (clusters) are less likely to be generalisable compared to studies where participants were randomly selected from numerous schools

throughout the country. In statistical terms the results of impact evaluations may present larger confidence intervals and larger standard errors. Randomised experiments, often report as sample sizes the number of learners contained in the programme and control groups as if it was done through a simple random sampling; however, this can be misleading if the cluster (number of schools or classes) is not properly taken into account in the calculation (Taylor, 2015). Table 30 illustrates how the standard errors of the effect size of some of the experimental studies reviewed in this meta-analysis can drastically change if cluster sampling is taken into account:

Table 30. *Change in Standard Error Once Adjusted by Sample Clusters*

		Ignoring clustering		**With clustering**	
Intervention	**Evaluation**	**Hedge's *g***	**Std. Error**	**Hedge's *g***	**Std. Error**
Khanya ICT	Louw et al. (2008)	0,09	0,087	0,09	0,161
RCUP	Fleisch et al. (2015)	0,03	0,041	0,03	0,100
B2B Workbooks	Schollar (2015)	0,64	0,038	0,64	0,155
Geography Study Guide	Taylor & Watson (2015)	0,14	0,019	0,14	0,060

For the above reason, the reviewers have decided to adjust the standard errors of the effect size in all the RCTs in this review based on cluster sampling principles in order to achieve more correct, precise and and fair estimates of results in this comparative meta-analysis.

5.7. Results of the statistical meta-analysis

5.7.1 Effect sizes at face-value

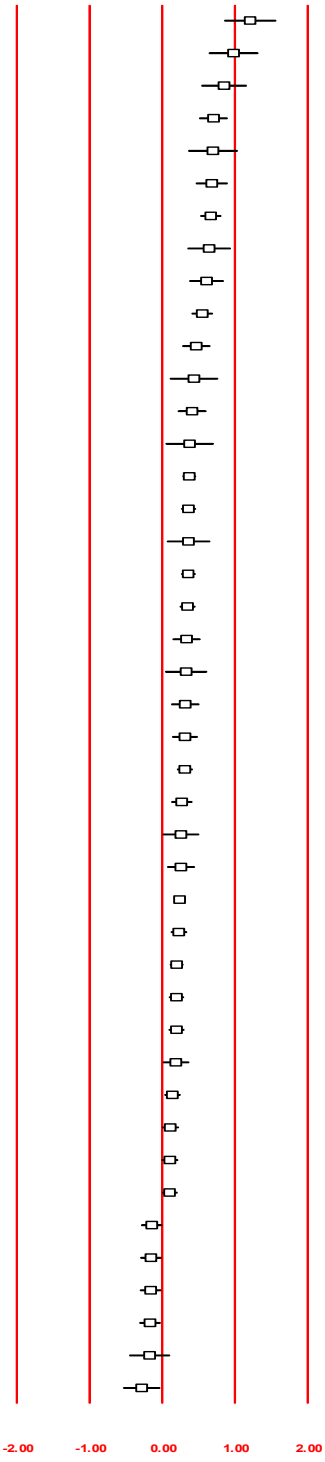
The overall results of the MTMA are summarised in the next general forest plot, which shows the magnitude of the effect size of each intervention on different learning outcomes. The 82 effects range from a lowest Hedge's *g* of -0,28 to a highest of +1,21. If an intervention had more than one effect size due to different learning outcomes or sub-groups being assessed, these were each calculated separately, and the cases would be identified for instance as Plus-Time A, Plus-Time B, Plus- Time C, and so forth.

The forest plot also illustrates the confidence interval of the effect sizes, which is reflective of the standard error, which is closely related to the sample sizes used in the study. The larger the standard error, the larger would also be the confidence range of the estimates. If the confidence interval of the effect size would include zero, the results would not be considered statistically significant. The forest plot below summarises the results of all the evaluations and impact studies included in this meta-analysis. The effect sizes which are not statistically significant or where the confidence interval intersects with zero have been excluded from Table 31.

As this is a comparative meta-analysis of different treatments with a lot of heterogeneity, the overall pooled results in the forest plot are of less interest to us at this stage.

Table 31. *Overview of Effect Sizes of SA education Interventions*

Intervention	Study	Learning Outcome	Hedges's g	Standard error
B2B Workbooks D (high dose)	Scholar (2013)	Math	1.210	0.185
B2B Workbooks C (high dose)	Scholar (2013)	Math	0.982	0.177
B2B Workbooks B	Scholar (2013)	Math	0.849	0.161
EQUIP F	Scholar (2002)	Math	0.705	0.103
E&O Math Challenge A	Scholar & Mouton (2014)	Math 30%	0.698	0.176
COUNT Family Math A	SAIDE (2007)	Math	0.683	0.115
GPLMS A	Fleisch & Schoer (2014)	English	0.669	0.076
B2B Workbooks A	Scholar (2013)	Math	0.645	0.155
Plus Time C	Prinsloo (2009)	Math	0.610	0.124
Dinaledi C	World Bank DIME (2010)	Science SG	0.550	0.077
EQUIP D CH5-7	Scholar (2002)	Math	0.468	0.102
E&O Math Challenge B	Scholar & Mouton (2014)	Math 50%	0.436	0.173
EQUIP B CH4-6	Scholar (2002)	Math	0.412	0.103
E&O Math Challenge D	Scholar & Mouton (2014)	Math Score Average	0.378	0.172
Learning for Living A	Scholar (2005)	English - Read	0.371	0.051
Learning for Living G CH3-7	Scholar (2005)	English - Read	0.362	0.056
B2B Workbooks E CH4-7	Scholar (2013)	Math	0.361	0.155
Learning for Living B	Scholar (2005)	English - Write	0.360	0.054
Learning for Living D	Scholar (2005)	English - Read	0.348	0.057
EQUIP C CH5-7	Scholar (2002)	English	0.334	0.101
Plus Time A	Prinsloo (2009)	English FAL	0.329	0.150
EQUIP E	Scholar (2002)	English	0.316	0.101
Numeric CAL	Bohmer (2014)	Math	0.313	0.092
Learning for Living F	Scholar (2005)	Math	0.311	0.058
Dinaledi D	World Bank DIME (2010)	Science HG	0.269	0.076
MPSI INSET F	Scholar (1999)	Math	0.257	0.131
EQUIP H	Scholar (2002)	Math	0.257	0.100
Language Companion	Botes & Mji (2010)	Math	0.239	0.041
Learning for Living E	Scholar (2005)	English - Write	0.229	0.060
Learning for Living C	Scholar (2005)	Math	0.199	0.051
Quality Learning Project C	Prinsloo & Kanjee (2005)	Math	0.196	0.056
Learning for Living H CH3-7	Scholar (2005)	English - Write	0.196	0.058
RCUP A	Fleisch, Taylor S, et al (2015)	Literacy - ANA	0.189	0.095
MTG Study Guides G	Taylor & Watson (2015)	Geography	0.141	0.060
Quality Learning Project A	Prinsloo & Kanjee (2005)	English	0.112	0.062
MTG Study Guides D	Taylor & Watson (2015)	Science - Life	0.106	0.060
Quality Learning Project D	Prinsloo & Kanjee (2005)	Math	0.104	0.057
Mine Proximity B	Besharati (2014)	Math SG	-0.144	0.078
Mine Proximity E	Besharati (2014)	Science SG	-0.154	0.078
Mine Proximity H	Besharati (2014)	NSC Batchelor	-0.160	0.078
Mine Proximity G	Besharati (2014)	NSC Pass	-0.168	0.078
B2B Workbooks G (high dose)	Fleisch, Taylor N et al (2011)	Math	-0.173	0.146
MPSI INSET D	Scholar (1999)	Math	-0.283	0.134



Prouced by authors (2015) using CMA3

5.7.2 Interpreting results in context

The fact that most studies above are statistically significant is a first good step. It tells us that the results are not just a phenomenon of chance. This is illustrated in the forest plot through small standard errors and by the fact that most of the confidence intervals of the impact estimates do not include zero as an element of the range.

While researchers are concerned with statistical significance, policy-makers are interested in educational significance and with the magnitude of the impact. This gives an indication to non-academic readers, policy-makers, investors, and the general public whether the interventions have a substantial and practical impact on real-life population outcomes. The question is whether the effect of an education interventions is large enough to warrant replication, transferability to public policy, extension and scale to the rest of the district, province or country? How do we know if an effect size is meaningful or not for education policy?

There are some general scales introduced by meta-analysis experts such as Cohen (1988) and Lipsey (1990) which provide a rough indication if effect sizes should be considered big or small.

Table 32. *'Rule of thumb' Interpretations of Effect Sizes*

Cohen's *d* (1988)	Lipsey's *d* (1990)	Interpretation
0,2	0,15	Small effect
0,4	0,45	Medium effect
0,8	0,90	Large effect

These guidelines are widely used in the social sciences, however they are merely a 'rule of thumb' (Table 32) and should not be used religiously. Hill et al. (2008) argue in fact that there are no universal guidelines for interpreting effect sizes. They suggest, rather, to develop empirical benchmarks which are contextualised in relationship to past research on the same type of interventions, population and outcome measures.

Effect sizes vary tremendously from sector to sector, and education programmes are known to have lower effect sizes than in other fields (McEwan, 2014; Valentine & Cooper, 2003). Different education analysts argue that interventions need to be taken seriously in the policy space, once they

demonstrate effect sizes over 0,25 (Coe, 2002), 0,4 (Hattie, 2003) and 0,59 (Marzano, 2004). What has also become clear is that different magnitudes of effect can be expected from different types of education interventions, with learner-targeted one-to-one tutoring, for example, producing much bigger effect sizes than whole school development programmes (Slavin, 2013).

The way evaluations are conducted and the methodologies that are used also matter greatly in meta-analysis. Lipsey et al. (2012) have for example discovered that impact studies using researcher self-developed testing instruments tend to yield higher effect sizes than evaluations which use broad standardised state-run assessments (such as for example the South African NSC exams). Whether RCTs or other quasi-experimental techniques are used, could also have an influence on the results achieved during an impact evaluation.

When conducting experiments in the education sector, Bloom et al. (2008) have also highlighted that effect magnitude may vary depending on the school levels in which the programmes have been implemented. The education and child developmental literature (Bloom et al., 2008, Hill et al., 2008) shows that even in the absence of an intervention a child is expected to improve its learning abilities year after year as a natural course of life. This is well illustrated in Table 33 that captures the average annual learning gains calculated from seven nationally normed assessments in the United States.

Table 33. *Normal Annual Learning Gains in American Schooling System*

Grade Transition	Reading Tests		Maths Tests	
	Average ES	Margin of Error	Average ES	Margin of Error
Grade K-1	1.52	0.21	1.14	0.49
Grade 1-2	0.97	0.10	1.03	0.14
Grade 2-3	0.60	0.10	0.89	0.16
Grade 3-4	0.36	0.12	0.52	0.14
Grade 4-5	0.40	0.06	0.56	0.11
Grade 5-6	0.32	0.11	0.41	0.08
Grade 6-7	0.23	0.11	0.30	0.06
Grade 7-8	0.26	0.03	0.32	0.05
Grade 8-9	0.24	0.10	0.22	0.10
Grade 9-10	0.19	0.08	0.25	0.07
Grade 10-11	0.19	0.17	0.14	0.16
Grade 11-12	0.06	0.11	0.01	0.14

Source: Hill et al. (2008)

Although these effect sizes may change in different countries, social contexts and with different subject outcomes being assessed, the trajectory of learning is clearly the same, where larger gains occur at lower grades and at younger ages while less learning occurs at older ages and at the end of the formal schooling process. This also highlights again the importance of using appropriate control groups and good counterfactual evaluation when assessing impact of interventions in the education sector, in order not to be misled by large learning gains occurring anyway in early grades of the schooling system (see for instance Fleisch et al., 2015)

Context, population, and learning outcomes all play a major role in affecting the results of impact studies and meta-analyses. The vast majority of field experiments and systematic reviews in the education sector are still conducted in the United States and in the industrialised world. Thus one of the best ways to benchmark the results from the South African impact evaluations is to assess the results against one another, as these were all implemented under similar circumstances and population context.

As discussed in section 2.9 of this research (see also Clemens et al., 2004; Harris, 2008), the time at which the impact of the intervention is measured can make a significant difference in the magnitude of the effect. Most interventions measure large effect sizes, during and shortly after they are implemented, due to the excitement and attention given to the beneficiaries during the project life-span (Schollar, 2013). But several years after the project has finished a natural process of 'decay' or 'compound' is expected to occur (Harris, 2008), therefore impact can look very different if measured in the short-term (right after the intervention), medium term or long-term (several years later). Almost all the effect sizes included in this meta-analysis have been measured at the end of the final year the intervention was administered. The only exception was in Schollar (2015), where the effect size of the PMRP workbooks was measured also 3 years later (in 2010), with effects of the programme still high though lower than when measured right after the treatment was provided in 2007.

From the preliminary results of this South African education meta-analysis we can see that the same intervention (PMRP workbooks) evaluated with the same technique (RCT) in two different studies yield strikingly opposite results, where the Back to the Basics Workbooks feature in the same meta-analysis (see forest plot above) as the top most impactful intervention as well as the third worst effective among the 74 effect sizes reviewed. This specific case will be discussed in more detail later in the chapter, but overall this goes to further prove that evaluation design and context does matter a lot, and effect sizes within a meta-analysis can be confounded by many other factors beyond the simple treatment intervention.

5.7.3 Exploring heterogeneity through moderator analysis

Having acknowledged that effect size is highly contingent on population and study design, it may not be appropriate to accept the results of the meta-analysis at face value. What would possibly be more useful is to explore in more depth the heterogeneity within these effects, so as to better understand which interventions seem to be most impactful within specific contexts and circumstances.

In the early stages of the review, all studies that met the inclusion criteria were subsequently coded one-by-one on the characteristics of the interventions, of the population, on the outcome measures and on the evaluation design. Each of these elements would potentially become 'predictor variables' or what is known in meta-analysis as 'moderators', which would be used to analyse heterogeneity in results. The various moderators would be statistically analysed to see if they had any influence on the effect sizes (as well as the standard errors) emerging from the studies.

To the extent possible, the reviewers tried to code as many moderators as possible in numerical format to allow for regression analysis. Some binary categorical variables could be converted into

dummy variables but others had to remain in categorical format and better analysed through ANOVA and sub-group analysis. Data was inputted, checked and organised first in Microsoft Excel and later brought into SPSS and CMA for more complex analysis. The variables coded as part of the review are illustrated in Table 34. The right column (Table 34) will provide an example from one study.

Table 34. *Study Coding Table*

Moderator	Category	Type	Example
Intervention	Case identifiers	Categorical	Amplats
Study		Categorical	Besharati (2014)
Effect Size (Hedge's *g*)	Outcome variables	Scale	-0.173
Standard Error		Scale	0.132
Type of intervention (Teacher-based, learner-based, LTSM, Whole School, etc.)	Intervention details	Categorical	Integrated School Development
Dosage of treatment (high / normal / low)		Categorical	n/a
Donor		Categorical	Anglo Platinum
Implementer		Categorical	Various
Lead institution (govern/private)		Dummy	Private
Duration of programme		Scale	5
School phase	Context	Ordinal	FET (4)
South African Province		Categorical	Limpopo & North West
Learning Outcome		Categorical	Science HG
Grade of Assessment		Scale	12

Table 34. *Study Coding Table (continued)*

Moderator	Category	Type	Example
Test Instrument		Categorical	NSC
Type of test instrument (evaluator's own / standardised state)	Evaluation / Study Details	Dummy	State
Type of measure		Categorical	Pass rate
Unit of analysis (school/learners)		Dummy	Schools
Date of baseline		Scale	Oct 2008
Date of post-test		Scale	Oct 2012
Years of exposure to treatment		Scale	4
Matched cohorts (same individuals pre-post / not)		Dummy	No
Counterfactual IE method		Categorical	PSM
Risk of bias		Ordinal	2
Sample size treated		Scale	90
Sample size control		Scale	159
Author		Categorical	Besharati
Year of publication		Scale	2014
Publisher		Categorical	SAIIA

Common tests of heterogeneity used in meta-analysis include *z*-test, Cochrane's Q (Walker, Hernandez, & Kattan, 2008), multivariate regression (Hedges, 1982; Raudenbush, 1994), chi square and I-Squared (Borenstein et al., 2009). This study utilised meta-regression, meta-ANOVA, and sub-group analysis as part of its approach to moderator analysis. The next sections will present and discuss the results of this exploration.

5.7.4 Meta-regression and meta-ANOVA

The first step in the moderator analysis chosen by the reviewer was to conduct a meta-regression to observe which of the many studies, interventions and contextual variables listed in Table 10 had a strong influence on the meta-analysis results (effect sizes and standard error). Previous education systematic reviews by McEwan (2014) which conducted meta-regression as part of moderator analysis have also highlighted that effect sizes may vary significantly once you control for other factors (simultaneous treatments, intervention features, context, methodology).

When conducting a meta-regression all the same assumptions, checks and rules which apply to a normal multivariate regression also need to be taken into consideration: large samples, linearity, normality, multi-collinearity, independence of cases, homoscedasticity. As explained previously the meta-analysis now contained 82 valid cases of effect sizes extracted from 18 high-quality studies of 20 interventions. If the meta-regression was conducted using the 'interventions' or 'studies' as the unit of analysis these would be too small a sample to yield enough statistical power to infer conclusions. Also the various effect sizes reported in each study were referring to different outcome measures and therefore would not be appropriate to collapse under one average effect size per study.

On the other hand, if the meta-regression is conducted at the level of 'effect sizes' (n=82) there would be more statistical power, however a lot of key assumptions of regression analysis would be violated. In our current meta-analysis one study cluster may yield anything from 1 to 22 effect sizes. This means that the effect sizes are actually 'nested' within the studies and therefore may suffer from the same standard errors, sampling, biases, quality concerns and variations caused by context and methodological design. The individual cases (the 82 effect sizes) used in the regression model are therefore not independent but in fact strongly correlated with other cases from the same study cluster.

Considering this challenge, the researcher decided to conduct a separate meta-regression using both 'effect sizes' and 'studies' as units of analysis. If results were consistent across the two levels, one could safely deduce that a certain moderator had a strong influence on the results of the meta-analysis. Considering the rule of thumb that the relationship between cases and variables in a regression should be roughly 1 to 10, the reviewers decided to test only 1 or 2 moderators at a time to explore if there was a significant and substantive correlation of the variable to the effect sizes. Meta-regression functions were available in Comprehensive Meta-Analysis version 3, which facilitated the analysis.

When doing the analysis with the 82 (non-independent) effect sizes, the study variables named 'type of test instrument', 'units of analysis', 'matched cohorts' and 'counterfactual IE method' appeared significantly correlated to the magnitude of Hedge's g. This however would not appear when the analysis was conducted at the level of studies (possibly due to low statistical power). The only variable that

yielded statistically significant correlation to the effect size in both levels of analysis was ‘grade of assessment’. The results were almost identical in both models. Table 35 presents the results from the regression conducted on the 82 effect sizes using ‘grade of assessment’ as the predictor variable.

Table 35. *Main Results for Regression Model, Random Effects (MM), z-Distribution, Hedges's g*

Covariate	Coefficient	Standard Error	Lower Bound	Upper Bound	Z-value	P-value
Intercept	0,431	0,064	0,306	0,557	6,73	0
Grade	-0,031	0,007	-0,045	-0,017	-4,34	0

Statistics for Regression Model

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

Q = 18.88, df = 1, p = 0.000

Goodness of fit: Test that unexplained variance is zero

Tau² = 0.03, Tau = 0.17, I² = 80.09%, Q = 401.89, df = 80, p = 0.000

Comparison of Regression Model with the null model

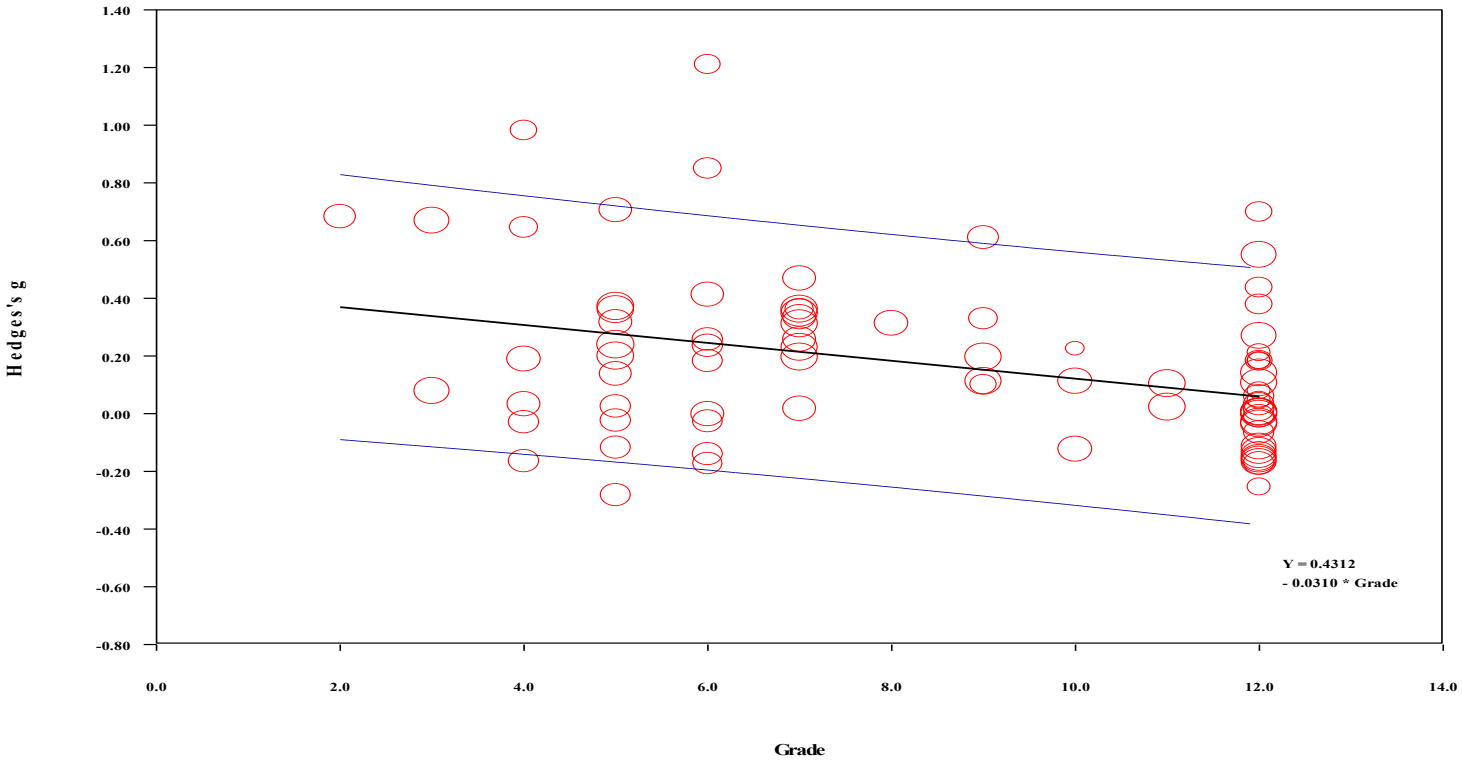
Total between-study variance (intercept only)

Tau² = 0.04, Tau = 0.21, I² = 85.06%, Q = 542.10, df = 81, p = 0.000

Proportion of total between-study variance explained by Regression Model

R² analog = 0.29

Figure 42. Regression of Hedge's *g* on Grade.



From the regression results (Table 35) we can observe that the grade in which the learning assessments are conducted has a clear negative correlation with the effect size, and this is also statistically significant. This explains at least some of the variances between the studies. This is also nicely illustrated in the above scatter plot of Hedge's *g* by grade of assessment, which shows that all studies regress along the same slope (black line), and are all contained within the same prediction interval (blue lines) (Figure 42). The only outliers are the effect sizes from the Schollar (2015) study, which had previously been flagged to produce exceptionally high results.

The finding from our meta-regression gives further proof to the general education theories and the meta-analytical consideration raised by Hill et al. (2008), that the higher the school grades in which the intervention occurs (and thus also where the learning assessments are conducted), the lower the effect sizes are to be expected. For the education policy debates, these findings further strengthen the call for more interventions to be conducted in the earlier years of schooling for more impact to be expected on learner achievement.

The meta-regression technique was useful to assess the correlation of continuous numerical variables on the outcome variable of effect size, however there are also some moderators which are dichotomous or categorical which are worth exploring using other techniques analogous to the analysis of variance (ANOVA). This is commonly conducted in meta-analysis as part of the analysis of heterogeneity.

We have seen earlier that certain parameters from the evaluation design could have potentially played a role in influencing the magnitude of the impact. The researcher decided therefore to test if there is a significant difference in the effect sizes of the various studies depending on the methodological framework that was used in the evaluation. As discussed earlier, there was an indication that 'counterfactual IE method', 'assessment instrument', 'units of analysis' and 'matched cohorts' could have influenced the magnitude of the impact results, a variance test was conducted of the methodological parameters used throughout the various studies. In a similar way to an ANOVA test, a heterogeneity test was performed using Q-value and TAU-squared statistics to analyse variance of different sub-groups of studies within the meta-analysis. The results are illustrated in Table 36.

Table 36. *Analysis of Heterogeneity Using Study Design Parameters*

		Effect size at 95% conf. interval						Heterogeneity				Tau-squared			
GROUPS	N (ES)	Mean	Std. error	Var.	Low. limit	Upp. limit	Z-value	Q-value	df (Q)	P-value	I-square	Tau Square	Std. Error	Var.	Tau
COUNTERFACTUAL IE METHOD (Mixed effects analysis)															
PSM	**20**	**0,003**	**0,024**	**0,001**	**-0,049**	**0,043**	**0,127**	**90,054**	**19**	**0,000**	**78,902**	**0,043**	**0,020**	**0,000**	**0,207**
RCT	**17**	**0,097**	**0,020**	**0,000**	**0,057**	**0,135**	**4.851**	**126.714**	**16**	**0,000**	**87,373**	**0,048**	**0,023**	**0,001**	**0,219**
RDD	**2**	**0,357**	**0,053**	**0,003**	**0,254**	**0,459**	**6.790**	**31,517**	**1**	**0,000**	**96,827**	**0,169**	**0,247**	**0,061**	**0,411**
simple match	**43**	**0,226**	**0,012**	**0,000**	**0,202**	**0,251**	**18.232**	**193.295**	**42**	**0,000**	**78,272**	**0,024**	**0,008**	**0,000**	**0,156**
Total between								**15,013**	**3**	**0,002**					
ASSESSMENT INSTRUMENT (Mixed effects analysis)															
own	**43**	**0,223**	**0,012**	**0,000**	**0,199**	**0,248**	**18.080**	**241,901**	**42**	**0,000**	**82,638**	**0,032**	**0,010**	**0,000**	**0,178**
state	**39**	**0,080**	**0,015**	**0,000**	**0,052**	**0,109**	**5.475**	**244,869**	**38**	**0,000**	**84,481**	**0,047**	**0,015**	**0,000**	**0,216**
Total between								**4.331**	**1**	**0,037**					

Table 36. *Analysis of Heterogeneity Using Study Design Parameters (continued)*

		Effect size at 95% conf. interval						Heterogeneity				Tau-squared			
GROUPS	N (ES)	Mean	Std. error	Var.	Low. limit	Upp. limit	Z-value	Q-value	df (Q)	P-value	I-square	Tau Square	Std. Error	Var.	Tau
UNITS OF ANALYSIS (Mixed effects analysis)															
Learners	**56**	**0,205**	**0,030**	**0,001**	**0,145**	**0,264**	**6,727**	**339.218**	**55**	**0,000**	**83,786**	**0,033**	**0,009**	**0,000**	**0,182**
Schools	**26**	**0,094**	**0.048**	**0,002**	**-0,000**	**0,188**	**1,968**	**181,639**	**25**	**0,000**	**86,236**	**0,073**	**0,028**	**0,001**	**0,269**
Total between								**3,830**	**1**	**0,050**					
MATCHED COHORTS OF LEARNERS (Mixed effects analysis)															
No	**50**	**0,138**	**0,012**	**0,000**	**0,115**	**0,162**	**11,637**	**343,793**	**49**	**0,000**	**85,747**	**0,043**	**0,012**	**0,000**	**0,208**
Yes	**32**	**0,208**	**0,016**	**0,000**	**0,178**	**0,239**	**13.373**	**185,578**	**31**	**0,000**	**83,295**	**0,040**	**0,016**	**0,000**	**0,200**
Total between								**6,083**	**1**	**0,014**					

What is evident from Table 36 is that all of the moderators of methodology (impact evaluation technique, units of analysis, test instrument, the use of pre- and post- matched cohorts), produced significantly different results within the meta-analysis, therefore a portion of the heterogeneity in the effect sizes is derived by the above variables and not just by the intervention itself. These results further confirm some of the findings of Lipsey et al. (2012) that impact studies generally report larger effects when researcher-developed test instruments (versus state assessments) are used to measure outcomes, and analysis is conducted at learner level rather than school level.

As demonstrated before in the meta-regression, the grade in which the learning assessments are conducted has a substantial and significant influence on the effect size. With the same logic, interventions conducted in lower phases of the South African school system should also report a larger impact than those occurring at the later stages of school grades. This assumption was tested again using the same ANOVA-like heterogeneity test and the results are illustrated in Table 37.

Table 37. *Analysis of Heterogeneity by School Phase*

GROUPS	N (ES)	Effect size at 95% conf. interval						Heterogeneity				Tau-squared			
		Mean	Std. error	Var.	Low. limit	Upp. limit	Z-value	Q-value	df (Q)	P-value	I-square	Tau Square	Std. Error	Var.	Tau
COUNTERFACTUAL IE METHOD (Mixed effects analysis)															
Prim - Found. (R-3)	**3**	**0,413**	**0,048**	**0,002**	**0,319**	**0,506**	**8.646**	**38,181**	**2**	**0,000**	**94,762**	**0,132**	**0,145**	**0,021**	**0,363**
Prim - Interm. (4-6)	**37**	**0,259**	**0,014**	**0,000**	**0,231**	**0,286**	**18.525**	**196,511**	**36**	**0,000**	**81,680**	**0,033**	**0,012**	**0,000**	**0,182**
Sec - Senior (7-9)	**6**	**0,222**	**0,035**	**0,001**	**0,154**	**0,290**	**6.403**	**15,034**	**5**	**0,010**	**66,742**	**0,017**	**0,018**	**0,000**	**0,129**
Sec - FET (10-12)	**36**	**0,030**	**0,014**	**0,00**	**0.000**	**0,059**	**2.092**	**130,609**	**35**	**0,000**	**73,202**	**0,021**	**0,008**	**0,000**	**0,145**
Total between								**25,205**	**3**	**0,000**					

Just as before with regression on grade of assessment, the pattern is unmistakable. The lower the phases of the school system in which interventions occur, the higher the impact observed on learning outcomes. Even with limited number of studies and effect sizes, the analysis above proves that the school phase is statistically and substantially significant in predicting the impact of the interventions.

As much as these meta-regressions provide some useful indications about the various factors that influence effect size, the results need to be interpreted with caution as they are based on a fairly limited number of cases which violate some of the key regression assumptions. As discussed earlier, the 82 effect sizes under analysis are not completely independent as they are correlated with one another within each study group. For this reason, a more appropriate analytical approach for this circumstance would be a hierarchical linear model (HLM) (Osborne, 2000; Raudenbush & Bryk, 2002; Woltman et al., 2012) which would regress 'Hedge's *g*' with various numerical or dichotomous parameters (education phase, assessment tool, units of analysis, matched cohorts) at the level-2 random effects, but keeping as fixed effects 'studies' at the level-1 of the model, where there is some shared variance among effect sizes. As CMA software did not have the capabilities for mixed level modelling, the data was transferred into SPSS. The 18 effect sizes in Besharati (2014) were further sub-divided into three other sub-groups linked to three different intervention evaluations (proximity to mine, Amplats whole school development programme, Radical M&S winter school).

Table 38. *HLM Regression with School Phase (Level 1), Studies/Interventions (Level 2)*

Estimates of Fixed Effects[a]							
						95% Confidence Interval	
Parameter	Estimate	Std. Error	df	t	Sig.	Lower Bound	Upper Bound
Intercept	.196314	.055638	17.634	3.528	.002	.079249	.313379
Phase	-.117099	.046258	47.441	-2.531	.015	-.210135	-.024062
Assessment	.072853	.136645	31.533	.533	.598	-.205646	.351352
Units	.011390	.161632	23.926	.070	.944	-.322258	.345038
Matched	.025670	.066601	74.839	.385	.701	-.107011	.158352

a. Dependent Variable: Hedges's *g*.

The multi-level regression model in Table 38 shows that, while the other parameters included in the model are not statistically significant, the school level in which the evaluation is conducted does still have a significant impact on the results. From the coefficient estimate one can also see that there is a negative correlation, where as the school phase increases the Hedge's *g* reported in the impact study decreases. As grade and school phase are such a prominent factor influencing effect size, from this point forward they will be included among the descriptive identifiers in the next sub-group analyses.

5.7.5 Sub-group analysis

Having established that heterogeneity in effect sizes is well expected in a MTMA and that this is not only a function of the intervention but also of context and study design, at this stage it is more useful to proceed with the analysis of the results using a sub-group approach. Rather than asking which is the most effective intervention to improve learning outcomes in South African schools, it is more appropriate to continue the investigation asking when, where, and how are education interventions effective?

The next few forest plots will break down the results of the meta-analysis by the different geographic contexts, learning areas, intervention types and implementing agencies. This will give some useful insights for policy-makers and investors who need to make decisions on which interventions to design, invest and support in different contexts of South Africa's education sector. Depending on the range of effects registered in each sub-group, forest plot scales have been adjusted from (-1/+1) to (-2/+2)

accordingly (for example with the inclusion of the Schollar 2015 results which include effect sizes above +1,0).

Table 39. *Interventions to Improve Language Abilities*.

Intervention	Study	Learning Outcome	Gr. Phase	Hedges's g	Standard error	Hedges's g and 95% CI
GPLMS A	Fleisch & Schoer (2014)	English	03 - Prim	0.669	0.076	
Learning for Living A	Schollar (2005)	English - Read	05 - Prim	0.371	0.051	
Learning for Living G CH3-7	Schollar (2005)	English - Read	07 - Prim	0.362	0.056	
Learning for Living B	Schollar (2005)	English - Write	05 - Prim	0.360	0.054	
Learning for Living D	Schollar (2005)	English - Read	07 - Prim	0.348	0.057	
EQUIP C CH5-7	Schollar (2002)	English	07 - Prim	0.334	0.101	
Plus Time A	Prinsloo (2009)	English FAL	09 - Sec	0.329	0.150	
EQUIP E	Schollar (2002)	English	05 - Prim	0.316	0.101	
Learning for Living E	Schollar (2005)	English - Write	07 - Prim	0.229	0.060	
Learning for Living H CH3-7	Schollar (2005)	English - Write	07 - Prim	0.196	0.058	
RCUP A	Fleisch, Taylor S, et al (2015)	Literacy - ANA	04 - Prim	0.189	0.095	
Quality Learning Project A	Prinsloo & Kanjee (2005)	English	09 - Sec	0.112	0.062	

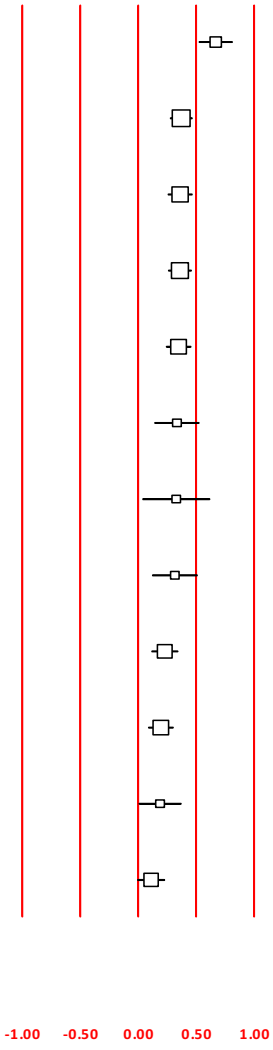
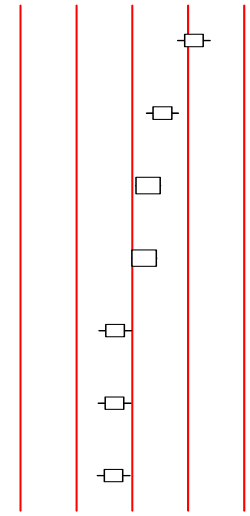


Table 40. *Interventions to Improve Science and Other FET Subjects*

Intervention	Study	Learning Outcome	Gr. Phase	Hedges's g	Standard error
Dinaledi C	World Bank DIME (2010)	Science SG	12 - Sec	0.550	0.077
Dinaledi D	World Bank DIME (2010)	Science HG	12 - Sec	0.269	0.076
MTG Study Guides G	Taylor & Watson (2015)	Geography	12 - Sec	0.141	0.060
MTG Study Guides D	Taylor & Watson (2015)	Science - Life	12 - Sec	0.106	0.060
Mine Proximity E	Besharati (2014)	Science SG	12 - Sec	-0.154	0.078
Mine Proximity H	Besharati (2014)	NSC Batchelor	12 - Sec	-0.160	0.078
Mine Proximity G	Besharati (2014)	NSC Pass	12 - Sec	-0.168	0.078



From the subject sub-group analysis (Table 39) one can observe that the literacy and numeracy interventions have a greater impact on learner achievement when implemented in primary schools such as in the case of the GPLMS and the Learning for Living, implemented in lower grades. Results from the Back 2 Basics workbooks yield extremely high results in the Schollar (2015) evaluation but very low results in the Fleisch et al. (2010) study.

As seen above, science and other subjects taught at higher stages of the schooling system, register a much lower impact than learning occurring in language and mathematics in lower grades. The Dinaledi (Table 40) however emerges as an exceptional case in which an intervention at FET level produces a substantial effect on science learning even when utilising grade-12 NSC exam results, which are known to normally yield relatively small effect sizes.

What seems to have a significant impact on high school pass rates and test scores is the introduction of additional teachers, assistants and human resources in order to improve educator-learner ratios, reduce class sizes, offer extra remedial classes and generally provide more time and attention to the individual learners, as illustrated by the findings emerging from the Besharati (2014) study on secondary school interventions in North West and Limpopo provinces, as well as from the evaluations of

the Dinaledi schools (Blum et al., 2010), Plus Time (Prinsloo, 2009) and the Epoch & Optima Trusts Maths Challenge Programme (Schollar & Mouton, 2014). This is also confirmed in the statistical analysis presented in the Tables 41-43.

The following forest plots will analyse the effect sizes of some of the most popular types of interventions frequently implemented in South Africa, such as the provision of LTSMs, teacher-based programmes and integrated school development.

Table 41. *Impact of Teacher-Based Interventions*

Intervention	Study	Learning Outcome	Gr. Phase	Hedges's g	Standard error
E&O Math Challenge A	Schollar & Mouton (2014)	Math 30%	12 - Sec	0.698	0.176
COUNT Family Math A	SAIDE (2007)	Math	02 - Prim	0.683	0.115
GPLMS A	Fleisch & Schoer (2014)	English	03 - Prim	0.669	0.076
Dinaledi C	World Bank DIME (2010)	Science SG	12 - Sec	0.550	0.077
E&O Math Challenge B	Schollar & Mouton (2014)	Math 50%	12 - Sec	0.436	0.173
E&O Math Challenge D	Schollar & Mouton (2014)	Math Score Average	12 - Sec	0.378	0.172
Dinaledi D	World Bank DIME (2010)	Science HG	12 - Sec	0.269	0.076
MPSI INSET F	Schollar (1999)	Math	06 - Prim	0.257	0.131
RCUP A	Fleisch, Taylor S, et al (2015)	Literacy - ANA	04 - Prim	0.189	0.095
MPSI INSET D	Schollar (1999)	Math	05 - Prim	-0.283	0.134
				0.411	0.033

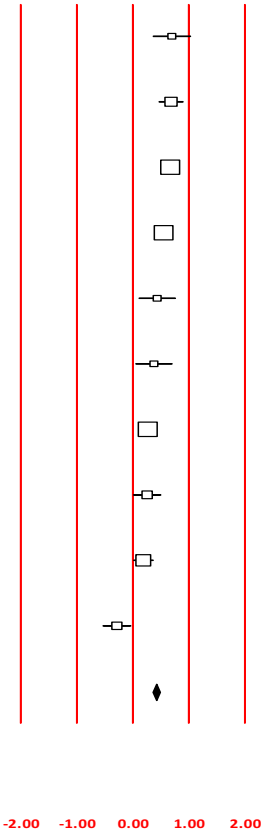


Table 42. *Impact of LTSM Provision*

Intervention	Study	Learning Outcome	Gr. Phase	Hedges's g	Standard error
B2B Workbooks D (high dose)	Schollar (2013)	Math	06 - Prim	1.210	0.185
B2B Workbooks C (high dose)	Schollar (2013)	Math	04 - Prim	0.982	0.177
B2B Workbooks B	Schollar (2013)	Math	06 - Prim	0.849	0.161
B2B Workbooks A	Schollar (2013)	Math	04 - Prim	0.645	0.155
B2B Workbooks E CH4-7	Schollar (2013)	Math	07 - Prim	0.361	0.155
Numeric CAL	Bohmer (2014)	Math	08 - Sec	0.313	0.092
Language Companion	Botes & Mji (2010)	Math	05 - Prim	0.239	0.041
MTG Study Guides G	Taylor & Watson (2015)	Geography	12 - Sec	0.141	0.060
MTG Study Guides D	Taylor & Watson (2015)	Science - Life	12 - Sec	0.106	0.060
B2B Workbooks G (high dose)	Fleisch, Taylor N et al (2011)	Math	06 - Prim	-0.173	0.146
				0.426	0.090

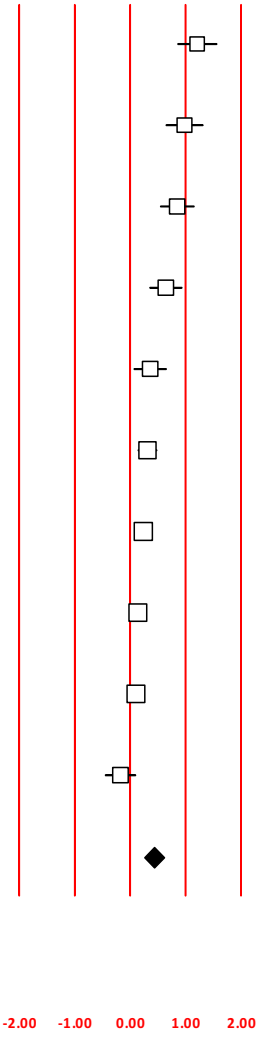
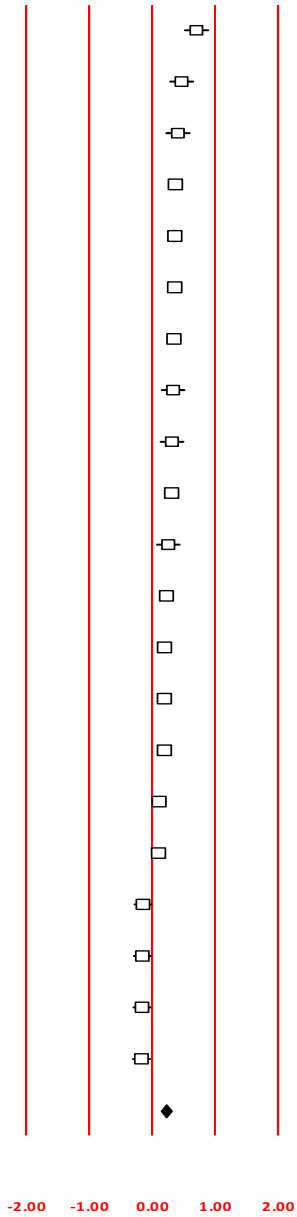


Table 43. *Impact of Integrated School Development Programmes*

Intervention	Study	Learning Outcome	Gr. Phase	Hedges's g	Standard error
EQUIP F	Schollar (2002)	Math	05 - Prim	0.705	0.103
EQUIP D CH5-7	Schollar (2002)	Math	07 - Prim	0.468	0.102
EQUIP B CH4-6	Schollar (2002)	Math	06 - Prim	0.412	0.103
Learning for Living A	Schollar (2005)	English - Read	05 - Prim	0.371	0.051
Learning for Living G CH3-7	Schollar (2005)	English - Read	07 - Prim	0.362	0.056
Learning for Living B	Schollar (2005)	English - Write	05 - Prim	0.360	0.054
Learning for Living D	Schollar (2005)	English - Read	07 - Prim	0.348	0.057
EQUIP C CH5-7	Schollar (2002)	English	07 - Prim	0.334	0.101
EQUIP E	Schollar (2002)	English	05 - Prim	0.316	0.101
Learning for Living F	Schollar (2005)	Math	07 - Prim	0.311	0.058
EQUIP H	Schollar (2002)	Math	07 - Prim	0.257	0.100
Learning for Living E	Schollar (2005)	English - Write	07 - Prim	0.229	0.060
Learning for Living C	Schollar (2005)	Math	05 - Prim	0.199	0.051
Quality Learning Project C	Prinsloo & Kanjee (2005)	Math	09 - Sec	0.196	0.056
Learning for Living H CH3-7	Schollar (2005)	English - Write	07 - Prim	0.196	0.058
Quality Learning Project A	Prinsloo & Kanjee (2005)	English	09 - Sec	0.112	0.062
Quality Learning Project D	Prinsloo & Kanjee (2005)	Math	11 - Sec	0.104	0.057
Mine Proximity B	Besharati (2014)	Math SG	12 - Sec	-0.144	0.078
Mine Proximity E	Besharati (2014)	Science SG	12 - Sec	-0.154	0.078
Mine Proximity H	Besharati (2014)	NSC Batchelor	12 - Sec	-0.160	0.078
Mine Proximity G	Besharati (2014)	NSC Pass	12 - Sec	-0.168	0.078
				0.218	0.043



Contrary to some of the international impact evaluations (Glewwe et al., 2007; Hanushek, 2002; Kremer et al., 2013) but in support to other education reviews (McEwan, 2014; Krishnaratne et al., 2013), evidence from South Africa shows that simple provision of effective learning and teaching material (i.e. Schollar, 2015; Taylor & Watson, 2015; Botes & Mji, 2010) can yield similar impact on learning

outcomes as the more complex and expensive whole school development programmes – shotgun approach (Mouton et al., 2014; Bloch, 2009). This finding, however, needs to be interpreted with caution as most programmes of LTSM provision in South Africa (i.e. READ, GPLMS, PMRP) are usually accompanied by teacher training, coaching and other school-level interventions.

Ironically again Back to the Basics (B2B) workbooks appear as the most effective as well as the least effective learning support material for grade 6 pupils in South Africa. Two well-performed RCTs by two different authors (Fleisch et al., 2010; Schollar, 2015) evaluating the same intervention which supposedly was implemented in the same manner, led one study to conclude that the workbooks have a phenomenal effect size of +1,21 while the other to report a disappointing effect size of -0,17. At first a reader will interpret this drastic difference in the results possibly caused by the different provincial contexts (Limpopo and Gauteng) but a closer analysis of the two studies will highlight that the variations lies in the methodological approach used in implementing the two experiments. Schollar used as his counterfactual 'null intervention' while Fleisch and Taylor utilised an alternative textbook (enhanced standardised practice), which arguably could have contained also some of the pedagogical elements of the B2B workbook. Eric Schollar and Associates (ESA) were very involved in the design and delivery of the intervention thus had a different set of interest and passion for the treatment than JET and Wits University evaluators, who also ended up using a different set of testing instruments from ESA.

Linked to the above case of the B2B workbooks, is the point that different intervention modalities are not necessarily mutually exclusive. Often the impact of one intervention is confounded by the effects of other treatment components (Banerjee et al., 2007; Bold et al., 2010; Friedman et al., 2010). For instance, many of the LTSMs provided by educational agencies (workbooks, lesson plans, science equipment, etc.) are often accompanied by systemic training of teachers and management on how to use them. Or the provision of extra teachers is also linked often to class and school re-structuring.

Many of the school development programmes discussed in this chapter contain very similar elements, methods and approaches to tackle learning. As such, Mouton (2014) argues that effectiveness evaluations of school development programmes conducted in a 'black-box' manner are less useful if they are not accompanied by an analysis of the processes in which change occurs. Deaton (2010) and later Schollar (2015) have argued that education evaluations should analyse and explore the 'mechanisms of change' that contribute to the improvement of learning outcomes in the broader schooling system rather than the specific education projects.

The next set of forest plots analyses the difference in effect sizes among the education interventions developed by government agencies, private sector and academic researchers (Tables 44 and 45). Even from an eye-ball glance it is clear (and somewhat expected), that interventions designed and implemented by academics in a semi-controlled environment, for the purpose of research,

experimentation and piloting, yield bigger effect sizes and more precise estimates than the on-going education policies and programmes run in large-scale policy environments by government or major private-public partnerships.

Table 44. *Impact of Government Programmes and Academic Experiments*

Group by Lead Agency	Intervention	Study	Learning Outcome	Gr. Phase	Hedges's g	Standard error
academic	B2B Workbooks D (high dose)	Schollar (2013)	Math	06 - Prim	1.210	0.185
academic	B2B Workbooks C (high dose)	Schollar (2013)	Math	04 - Prim	0.982	0.177
academic	B2B Workbooks B	Schollar (2013)	Math	06 - Prim	0.849	0.161
academic	B2B Workbooks A	Schollar (2013)	Math	04 - Prim	0.645	0.155
academic	B2B Workbooks E CH4-7	Schollar (2013)	Math	07 - Prim	0.361	0.155
academic	Numeric CAL	Bohmer (2014)	Math	08 - Sec	0.313	0.092
academic	Language Companion	Botes & Mji (2010)	Math	05 - Prim	0.239	0.041
academic					0.629	0.133
government	GPLMS A	Fleisch & Schoer (2014)	English	03 - Prim	0.669	0.076
government	Plus Time C	Prinsloo (2009)	Math	09 - Sec	0.610	0.124
government	Dinaledi C	World Bank DIME (2010)	Science SG	12 - Sec	0.550	0.077
government	Dinaledi D	World Bank DIME (2010)	Science HG	12 - Sec	0.269	0.076
government	MPSI INSET F	Schollar (1999)	Math	06 - Prim	0.257	0.131
government	MTG Study Guides G	Taylor & Watson (2015)	Geography	12 - Sec	0.141	0.060
government	MTG Study Guides D	Taylor & Watson (2015)	Science - Life	12 - Sec	0.106	0.060
government	MPSI INSET D	Schollar (1999)	Math	05 - Prim	-0.283	0.134
government					0.295	0.097

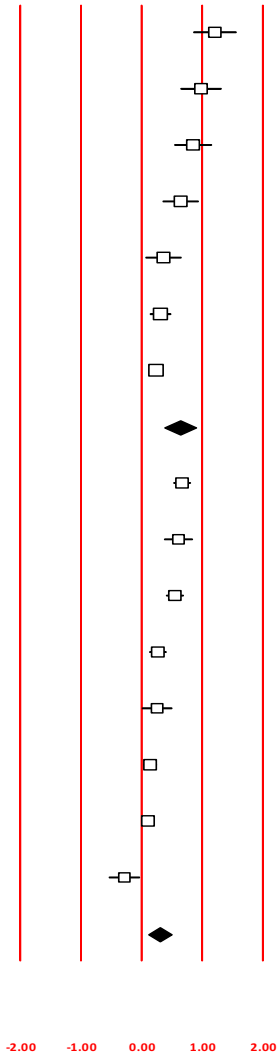
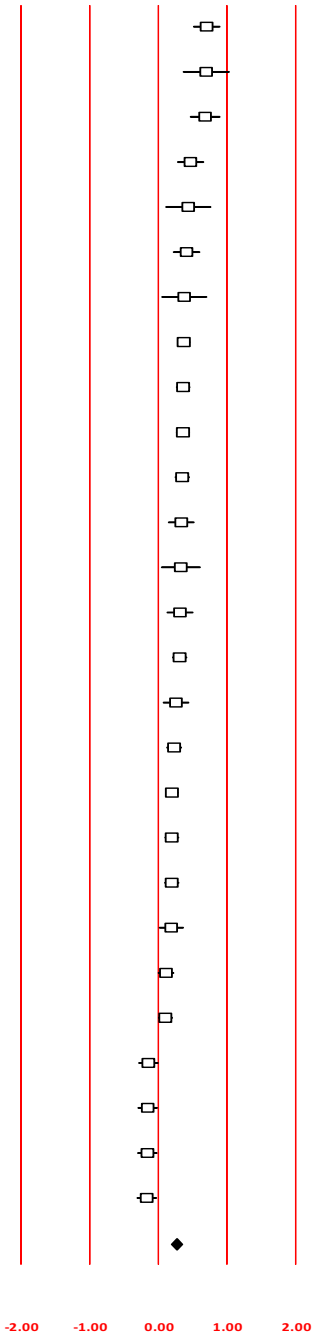


Table 45. *Impact of Education Programmes Led By Private Sector*

Intervention	Study	Learning Outcome	Gr. Phase	Hedges's g	Standard error
EQUIP F	Schollar (2002)	Math	05 - Prim	0.705	0.103
E&O Math Challenge A	Schollar & Mouton (2014)	Math 30%	12 - Sec	0.698	0.176
COUNT Family Math A	SAIDE (2007)	Math	02 - Prim	0.683	0.115
EQUIP D CH5-7	Schollar (2002)	Math	07 - Prim	0.468	0.102
E&O Math Challenge B	Schollar & Mouton (2014)	Math 50%	12 - Sec	0.436	0.173
EQUIP B CH4-6	Schollar (2002)	Math	06 - Prim	0.412	0.103
E&O Math Challenge D	Schollar & Mouton (2014)	Math Score Average	12 - Sec	0.378	0.172
Learning for Living A	Schollar (2005)	English - Read	05 - Prim	0.371	0.051
Learning for Living G CH3-7	Schollar (2005)	English - Read	07 - Prim	0.362	0.056
Learning for Living B	Schollar (2005)	English - Write	05 - Prim	0.360	0.054
Learning for Living D	Schollar (2005)	English - Read	07 - Prim	0.348	0.057
EQUIP C CH5-7	Schollar (2002)	English	07 - Prim	0.334	0.101
Plus Time A	Prinsloo (2009)	English FAL	09 - Sec	0.329	0.150
EQUIP E	Schollar (2002)	English	05 - Prim	0.316	0.101
Learning for Living F	Schollar (2005)	Math	07 - Prim	0.311	0.058
EQUIP H	Schollar (2002)	Math	07 - Prim	0.257	0.100
Learning for Living E	Schollar (2005)	English - Write	07 - Prim	0.229	0.060
Learning for Living C	Schollar (2005)	Math	05 - Prim	0.199	0.051
Quality Learning Project C	Prinsloo & Kanjee (2005)	Math	09 - Sec	0.196	0.056
Learning for Living H CH3-7	Schollar (2005)	English - Write	07 - Prim	0.196	0.058
RCUP A	Fleisch, Taylor S, et al (2015)	Literacy - ANA	04 - Prim	0.189	0.095
Quality Learning Project A	Prinsloo & Kanjee (2005)	English	09 - Sec	0.112	0.062
Quality Learning Project D	Prinsloo & Kanjee (2005)	Math	11 - Sec	0.104	0.057
Mine Proximity B	Besharati (2014)	Math SG	12 - Sec	-0.144	0.078
Mine Proximity E	Besharati (2014)	Science SG	12 - Sec	-0.154	0.078
Mine Proximity H	Besharati (2014)	NSC Batchelor	12 - Sec	-0.160	0.078
Mine Proximity G	Besharati (2014)	NSC Pass	12 - Sec	-0.168	0.078
				0.256	0.040



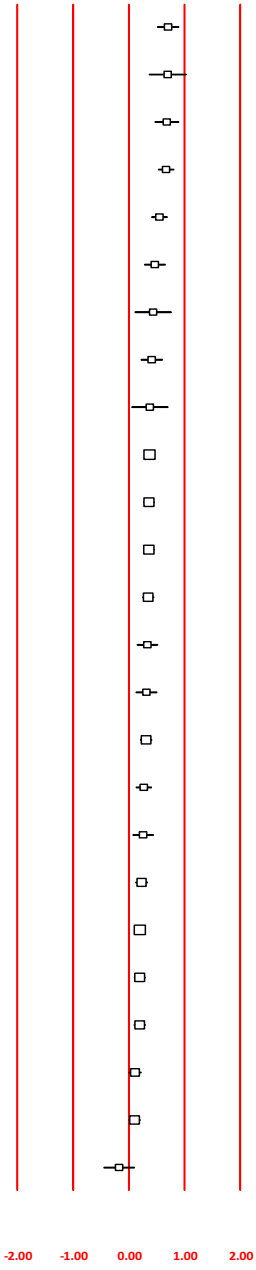
Prouced by authors (2015) using CMA3

The last set of forest plots will review the impact of interventions in six South African provinces where many school development programmes have been implemented over the course of the past 15 years (Tables 46 -51).

Although various studies of South Africa's education sector (i.e., Fleisch and Schoer, 2014; Gustaffson and Taylor, 2013; Van der Berg, 2008; World Bank, 2010), suggest that programmes have different impacts in different geographic areas and socio-economic contexts, the results from the above meta-analysis illustrate that geographic locations within South Africa seem to provide relatively little heterogeneity with regard to the impact of the interventions, compared to other influencing factors such as education phase and evaluation design. Nonetheless the above geographic sub-group analysis can be useful to the provincial education authorities and private investors when reviewing the interventions, which have worked best in their specific areas of jurisdiction and region of focus (Tables 46-51).

Table 46. *Impact of Education Interventions in Gauteng*

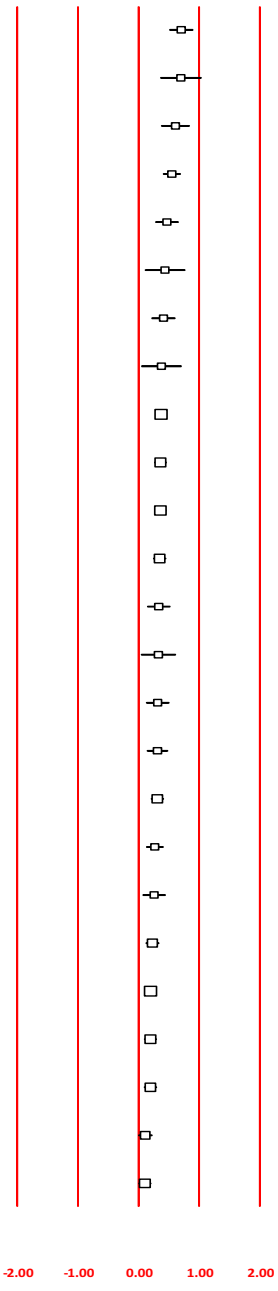
Intervention	Study	Learning Outcome	Gr. Phase	Hedges's g	Standard error	Hedges's g and 95% CI
EQUIP F	Schollar (2002)	Math	05 - Prim	0.705	0.103	
E&O Math Challenge A	Schollar & Mouton (2014)	Math 30%	12 - Sec	0.698	0.176	
COUNT Family Math A	SAIDE (2007)	Math	02 - Prim	0.683	0.115	
GPLMS A	Fleisch & Schoer (2014)	English	03 - Prim	0.669	0.076	
Dinaledi C	World Bank DIME (2010)	Science SG	12 - Sec	0.550	0.077	
EQUIP D CH5-7	Schollar (2002)	Math	07 - Prim	0.468	0.102	
E&O Math Challenge B	Schollar & Mouton (2014)	Math 50%	12 - Sec	0.436	0.173	
EQUIP B CH4-6	Schollar (2002)	Math	06 - Prim	0.412	0.103	
E&O Math Challenge D	Schollar & Mouton (2014)	Math Score Average	12 - Sec	0.378	0.172	
Learning for Living A	Schollar (2005)	English - Read	05 - Prim	0.371	0.051	
Learning for Living G CH3-7	Schollar (2005)	English - Read	07 - Prim	0.362	0.056	
Learning for Living B	Schollar (2005)	English - Write	05 - Prim	0.360	0.054	
Learning for Living D	Schollar (2005)	English - Read	07 - Prim	0.348	0.057	
EQUIP C CH5-7	Schollar (2002)	English	07 - Prim	0.334	0.101	
EQUIP E	Schollar (2002)	English	05 - Prim	0.316	0.101	
Learning for Living F	Schollar (2005)	Math	07 - Prim	0.311	0.058	
Dinaledi D	World Bank DIME (2010)	Science HG	12 - Sec	0.269	0.076	
EQUIP H	Schollar (2002)	Math	07 - Prim	0.257	0.100	
Learning for Living E	Schollar (2005)	English - Write	07 - Prim	0.229	0.060	
Learning for Living C	Schollar (2005)	Math	05 - Prim	0.199	0.051	
Quality Learning Project C	Prinsloo & Kanjee (2005)	Math	09 - Sec	0.196	0.056	
Learning for Living H CH3-7	Schollar (2005)	English - Write	07 - Prim	0.196	0.058	
Quality Learning Project A	Prinsloo & Kanjee (2005)	English	09 - Sec	0.112	0.062	
Quality Learning Project D	Prinsloo & Kanjee (2005)	Math	11 - Sec	0.104	0.057	
B2B Workbooks G (high dose)	Fleisch, Taylor N et al (2011)	Math	06 - Prim	-0.173	0.146	



Prouced by authors (2015) using CMA3

Table 47. *Impact of Education Interventions in Western Cape.*

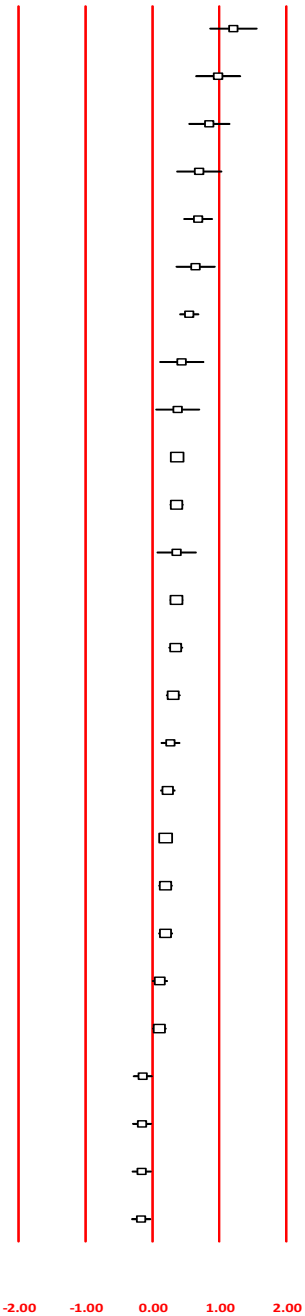
Intervention	Study	Learning Outcome	Gr. Phase	Hedges's g	Standard error	Hedges's g and 95% CI
EQUIP F	Schollar (2002)	Math	05 - Prim	0.705	0.103	
E&O Math Challenge A	Schollar & Mouton (2014)	Math 30%	12 - Sec	0.698	0.176	
Plus Time C	Prinsloo (2009)	Math	09 - Sec	0.610	0.124	
Dinaledi C	World Bank DIME (2010)	Science SG	12 - Sec	0.550	0.077	
EQUIP D CH5-7	Schollar (2002)	Math	07 - Prim	0.468	0.102	
E&O Math Challenge B	Schollar & Mouton (2014)	Math 50%	12 - Sec	0.436	0.173	
EQUIP B CH4-6	Schollar (2002)	Math	06 - Prim	0.412	0.103	
E&O Math Challenge D	Schollar & Mouton (2014)	Math Score Average	12 - Sec	0.378	0.172	
Learning for Living A	Schollar (2005)	English - Read	05 - Prim	0.371	0.051	
Learning for Living G CH3-7	Schollar (2005)	English - Read	07 - Prim	0.362	0.056	
Learning for Living B	Schollar (2005)	English - Write	05 - Prim	0.360	0.054	
Learning for Living D	Schollar (2005)	English - Read	07 - Prim	0.348	0.057	
EQUIP C CH5-7	Schollar (2002)	English	07 - Prim	0.334	0.101	
Plus Time A	Prinsloo (2009)	English FAL	09 - Sec	0.329	0.150	
EQUIP E	Schollar (2002)	English	05 - Prim	0.316	0.101	
Numeric CAL	Bohmer (2014)	Math	08 - Sec	0.313	0.092	
Learning for Living F	Schollar (2005)	Math	07 - Prim	0.311	0.058	
Dinaledi D	World Bank DIME (2010)	Science HG	12 - Sec	0.269	0.076	
EQUIP H	Schollar (2002)	Math	07 - Prim	0.257	0.100	
Learning for Living E	Schollar (2005)	English - Write	07 - Prim	0.229	0.060	
Learning for Living C	Schollar (2005)	Math	05 - Prim	0.199	0.051	
Quality Learning Project C	Prinsloo & Kanjee (2005)	Math	09 - Sec	0.196	0.056	
Learning for Living H CH3-7	Schollar (2005)	English - Write	07 - Prim	0.196	0.058	
Quality Learning Project A	Prinsloo & Kanjee (2005)	English	09 - Sec	0.112	0.062	
Quality Learning Project D	Prinsloo & Kanjee (2005)	Math	11 - Sec	0.104	0.057	



Prouced by authors (2015) using CMA3

Table 48. *Impact of Education Interventions in Limpopo.*

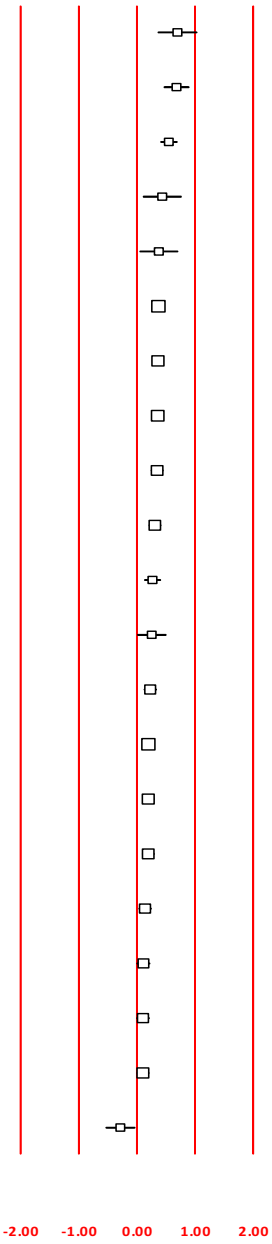
Intervention	Study	Learning Outcome	Gr. Phase	Hedges's g	Standard error
B2B Workbooks D (high dose)	Schollar (2013)	Math	06 - Prim	1.210	0.185
B2B Workbooks C (high dose)	Schollar (2013)	Math	04 - Prim	0.982	0.177
B2B Workbooks B	Schollar (2013)	Math	06 - Prim	0.849	0.161
E&O Math Challenge A	Schollar & Mouton (2014)	Math 30%	12 - Sec	0.698	0.176
COUNT Family Math A	SAIDE (2007)	Math	02 - Prim	0.683	0.115
B2B Workbooks A	Schollar (2013)	Math	04 - Prim	0.645	0.155
Dinaledi C	World Bank DIME (2010)	Science SG	12 - Sec	0.550	0.077
E&O Math Challenge B	Schollar & Mouton (2014)	Math 50%	12 - Sec	0.436	0.173
E&O Math Challenge D	Schollar & Mouton (2014)	Math Score Average	12 - Sec	0.378	0.172
Learning for Living A	Schollar (2005)	English - Read	05 - Prim	0.371	0.051
Learning for Living G CH3-7	Schollar (2005)	English - Read	07 - Prim	0.362	0.056
B2B Workbooks E CH4-7	Schollar (2013)	Math	07 - Prim	0.361	0.155
Learning for Living B	Schollar (2005)	English - Write	05 - Prim	0.360	0.054
Learning for Living D	Schollar (2005)	English - Read	07 - Prim	0.348	0.057
Learning for Living F	Schollar (2005)	Math	07 - Prim	0.311	0.058
Dinaledi D	World Bank DIME (2010)	Science HG	12 - Sec	0.269	0.076
Learning for Living E	Schollar (2005)	English - Write	07 - Prim	0.229	0.060
Learning for Living C	Schollar (2005)	Math	05 - Prim	0.199	0.051
Quality Learning Project C	Prinsloo & Kanjee (2005)	Math	09 - Sec	0.196	0.056
Learning for Living H CH3-7	Schollar (2005)	English - Write	07 - Prim	0.196	0.058
Quality Learning Project A	Prinsloo & Kanjee (2005)	English	09 - Sec	0.112	0.062
Quality Learning Project D	Prinsloo & Kanjee (2005)	Math	11 - Sec	0.104	0.057
Mine Proximity B	Besharati (2014)	Math SG	12 - Sec	-0.144	0.078
Mine Proximity E	Besharati (2014)	Science SG	12 - Sec	-0.154	0.078
Mine Proximity H	Besharati (2014)	NSC Batchelor	12 - Sec	-0.160	0.078
Mine Proximity G	Besharati (2014)	NSC Pass	12 - Sec	-0.168	0.078



Prouced by authors (2015) using CMA3

Table 49. *Impact of Education Interventions in KwaZulu Natal*

Intervention	Study	Learning Outcome	Gr. Phase	Hedges's g	Standard error	Hedges's g and 95% CI
E&O Math Challenge A	Schollar & Mouton (2014)	Math 30%	12 - Sec	0.698	0.176	
COUNT Family Math A	SAIDE (2007)	Math	02 - Prim	0.683	0.115	
Dinaledi C	World Bank DIME (2010)	Science SG	12 - Sec	0.550	0.077	
E&O Math Challenge B	Schollar & Mouton (2014)	Math 50%	12 - Sec	0.436	0.173	
E&O Math Challenge D	Schollar & Mouton (2014)	Math Score Average	12 - Sec	0.378	0.172	
Learning for Living A	Schollar (2005)	English - Read	05 - Prim	0.371	0.051	
Learning for Living G CH3-7	Schollar (2005)	English - Read	07 - Prim	0.362	0.056	
Learning for Living B	Schollar (2005)	English - Write	05 - Prim	0.360	0.054	
Learning for Living D	Schollar (2005)	English - Read	07 - Prim	0.348	0.057	
Learning for Living F	Schollar (2005)	Math	07 - Prim	0.311	0.058	
Dinaledi D	World Bank DIME (2010)	Science HG	12 - Sec	0.269	0.076	
MPSI INSET F	Schollar (1999)	Math	06 - Prim	0.257	0.131	
Learning for Living E	Schollar (2005)	English - Write	07 - Prim	0.229	0.060	
Learning for Living C	Schollar (2005)	Math	05 - Prim	0.199	0.051	
Quality Learning Project C	Prinsloo & Kanjee (2005)	Math	09 - Sec	0.196	0.056	
Learning for Living H CH3-7	Schollar (2005)	English - Write	07 - Prim	0.196	0.058	
MTG Study Guides G	Taylor & Watson (2015)	Geography	12 - Sec	0.141	0.060	
Quality Learning Project A	Prinsloo & Kanjee (2005)	English	09 - Sec	0.112	0.062	
MTG Study Guides D	Taylor & Watson (2015)	Science - Life	12 - Sec	0.106	0.060	
Quality Learning Project D	Prinsloo & Kanjee (2005)	Math	11 - Sec	0.104	0.057	
MPSI INSET D	Schollar (1999)	Math	05 - Prim	-0.283	0.134	



Prouced by authors (2015) using CMA3

Table 50. *Impact of Education Interventions In Mpumalanga*

Intervention	Study	Learning Outcome	Gr. Phase	Hedges's g	Standard error	Hedges's g and 95% CI
E&O Math Challenge A	Schollar & Mouton (2014)	Math 30%	12 - Sec	0.698	0.176	
COUNT Family Math A	SAIDE (2007)	Math	02 - Prim	0.683	0.115	
Dinaledi C	World Bank DIME (2010)	Science SG	12 - Sec	0.550	0.077	
E&O Math Challenge B	Schollar & Mouton (2014)	Math 50%	12 - Sec	0.436	0.173	
E&O Math Challenge D	Schollar & Mouton (2014)	Math Score Average	12 - Sec	0.378	0.172	
Learning for Living A	Schollar (2005)	English - Read	05 - Prim	0.371	0.051	
Learning for Living G CH3-7	Schollar (2005)	English - Read	07 - Prim	0.362	0.056	
Learning for Living B	Schollar (2005)	English - Write	05 - Prim	0.360	0.054	
Learning for Living D	Schollar (2005)	English - Read	07 - Prim	0.348	0.057	
Learning for Living F	Schollar (2005)	Math	07 - Prim	0.311	0.058	
Dinaledi D	World Bank DIME (2010)	Science HG	12 - Sec	0.269	0.076	
MPSI INSET F	Schollar (1999)	Math	06 - Prim	0.257	0.131	
Learning for Living E	Schollar (2005)	English - Write	07 - Prim	0.229	0.060	
Learning for Living C	Schollar (2005)	Math	05 - Prim	0.199	0.051	
Quality Learning Project C	Prinsloo & Kanjee (2005)	Math	09 - Sec	0.196	0.056	
Learning for Living H CH3-7	Schollar (2005)	English - Write	07 - Prim	0.196	0.058	
MTG Study Guides G	Taylor & Watson (2015)	Geography	12 - Sec	0.141	0.060	
Quality Learning Project A	Prinsloo & Kanjee (2005)	English	09 - Sec	0.112	0.062	
MTG Study Guides D	Taylor & Watson (2015)	Science - Life	12 - Sec	0.106	0.060	
Quality Learning Project D	Prinsloo & Kanjee (2005)	Math	11 - Sec	0.104	0.057	
MPSI INSET D	Schollar (1999)	Math	05 - Prim	-0.283	0.134	

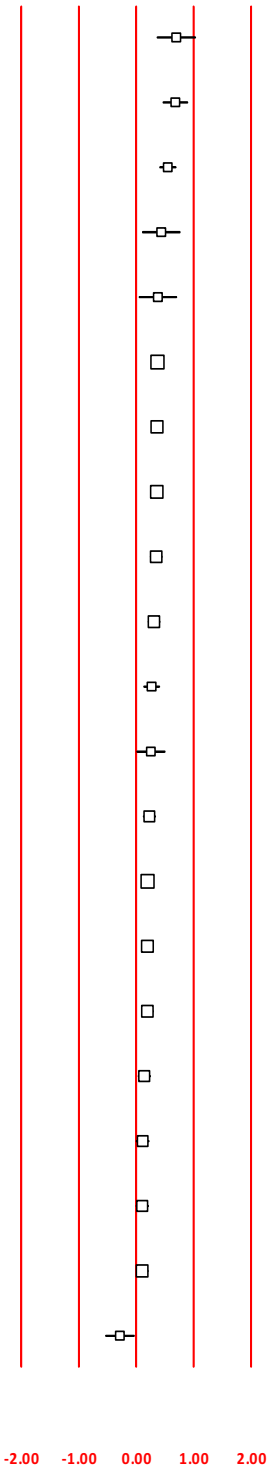
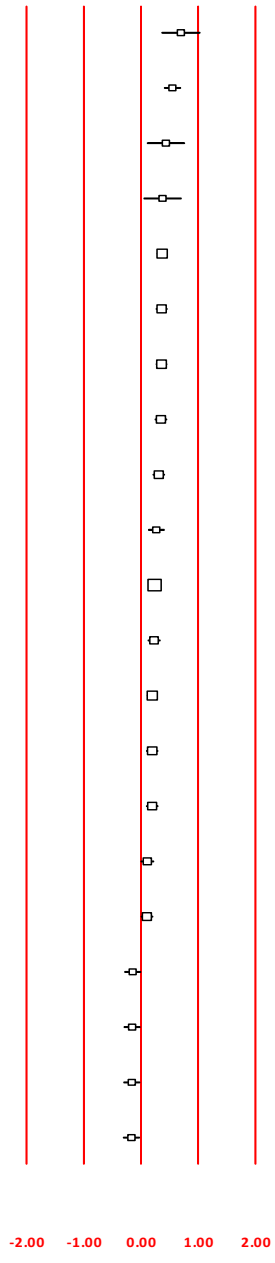


Table 51. *Impact of Education Interventions In North West*

Intervention	Study	Learning Outcome	Gr. Phase	Hedges's g	Standard error	Hedges's g and 95% CI
E&O Math Challenge A	Schollar & Mouton (2014)	Math 30%	12 - Sec	0.698	0.176	
Dinaledi C	World Bank DIME (2010)	Science SG	12 - Sec	0.550	0.077	
E&O Math Challenge B	Schollar & Mouton (2014)	Math 50%	12 - Sec	0.436	0.173	
E&O Math Challenge D	Schollar & Mouton (2014)	Math Score Average	12 - Sec	0.378	0.172	
Learning for Living A	Schollar (2005)	English - Read	05 - Prim	0.371	0.051	
Learning for Living G CH3-7	Schollar (2005)	English - Read	07 - Prim	0.362	0.056	
Learning for Living B	Schollar (2005)	English - Write	05 - Prim	0.360	0.054	
Learning for Living D	Schollar (2005)	English - Read	07 - Prim	0.348	0.057	
Learning for Living F	Schollar (2005)	Math	07 - Prim	0.311	0.058	
Dinaledi D	World Bank DIME (2010)	Science HG	12 - Sec	0.269	0.076	
Language Companion	Botes & Mji (2010)	Math	05 - Prim	0.239	0.041	
Learning for Living E	Schollar (2005)	English - Write	07 - Prim	0.229	0.060	
Learning for Living C	Schollar (2005)	Math	05 - Prim	0.199	0.051	
Quality Learning Project C	Prinsloo & Kanjee (2005)	Math	09 - Sec	0.196	0.056	
Learning for Living H CH3-7	Schollar (2005)	English - Write	07 - Prim	0.196	0.058	
Quality Learning Project A	Prinsloo & Kanjee (2005)	English	09 - Sec	0.112	0.062	
Quality Learning Project D	Prinsloo & Kanjee (2005)	Math	11 - Sec	0.104	0.057	
Mine Proximity B	Besharati (2014)	Math SG	12 - Sec	-0.144	0.078	
Mine Proximity E	Besharati (2014)	Science SG	12 - Sec	-0.154	0.078	
Mine Proximity H	Besharati (2014)	NSC Batchelor	12 - Sec	-0.160	0.078	
Mine Proximity G	Besharati (2014)	NSC Pass	12 - Sec	-0.168	0.078	



Prouced by authors (2015) using CMA3

5.8. Failure of cost-effectiveness comparison

Accounting for costs during a comparative meta-analysis becomes an even more useful exercise to assist policy-makers and investors than simply estimating programme impact. The analytical framework of this research thus makes provision for the integration of a CER as part of the meta-analysis and the additional forest plots being developed.

As the CER is constructed using effect sizes, which might range in sampling errors, CER estimates will also be subject to the same variances, standard errors and confidence intervals. The cost values are usually also just estimates and subject to a confidence range, therefore CER presents even less conclusive evidence than simple effect sizes. Not all effects are positive, substantial or statistically significant, therefore for practical purposes it would be reasonable to calculate CER only for the interventions with large and significant effect sizes, whose lower confidence bound is above the zero line.

Integrating cost-effectiveness in this South African education meta-analysis research, proved to be extremely difficult. The first challenge was in retrieving budgetary information about the various interventions. Because of political and commercial sensitivities, both public and private institutions were not forthcoming in disclosing their accounts to external parties.

When information was not available in the official programme reports, the reviewers contacted the original evaluators, and at times even the managers and donors of the respective programmes (Table 52). Generally, the donors, who had a keen interest in knowing the impact of their investments, were the most forthcoming and reliable source of cost information.

Table 52. *Interventions For Which Some Cost Information Was Retrieved*

Programme Name	Source of information
Amplats Education Programme	Anglo American (donor)
Radical M&S Winter School	Anglo American (donor)
Quality of Learning Programme	Business Trust (donor)
Learning for Living	Business Trust (donor)

Table 52. *Interventions For Which Some Cost Information Was Retrieved (continued)*

Programme Name	Source of information
Plus Time	Shuttleworth Foundation (donor)
Primary Maths Research Project	Eric Schollar (developer & evaluator)
MTG Study Guides	Stephen Taylor (researcher)
Problem-based learning teacher training	Annemarie Hattingh (researcher)

Where financial information was made available this was often unclear, inaccurate, incomplete or the information was not presented in a consistent manner across the various organisational/programme reports. A large portion of the cost information was only estimates, and expressed as budgetary commitments rather than actual disbursements or expenditures. Taking the total lump-sum budget of programmes at face value without disaggregating in detail all the costs. Without understanding the full depth, length and scope of what was included in the budget, would result in very misleading CER results.

Some budgets reviewed included major capital costs (i.e. construction of school facilities, provision of LTSM), which had a life span far beyond the project. Some interventions were made possible also thanks to in-kind donations (free goods or voluntary services) which were often not reflected in budgets. Any intervention would require a certain degree of administrative support and managerial oversight costs, which was often shared between projects, and therefore difficult to properly quantify. Untying the specific intervention costs from the broader system costs (such as public official's time on project) was particularly challenging especially when costing of government programmes such as for instance the GPLMS, CTLI, MTG study guides, etc.

Most private sector projects would include the costs to the donor and implementing organisation, but not the user (learners, families, teachers), such as the opportunity and forgone costs for the participants to attend the specific programmes, which sometimes could be quite taxing. Different projects demanded different degrees of engagement of national and provincial government and of civil networks that were usually not accounted for in the project budgets. Some of the big country-wide programmes

such as the Dinaledi, the EQUIP, the QLP, had the additional challenge of being implemented by a range of different service providers who used different approaches, dosages and qualities of goods and services, making it very difficult to standardise the costs. This became even more difficult when having to put a monetary value to the quality of different kind of teachers, trainers and specialised technical assistance provided.

The evaluation design also created further complications for the CEA. Effect sizes were normally computed on specific learning outcomes, but most of the time programmes were designed to address multiple outcomes at the same time (for instance language learning, mathematics learning, as well as general school attendance). The challenge was then how to sub-divide the programme costs by the various outcomes the programme tried to reach. Also effect sizes were calculated using a specific sample of programme participants, while organisational reports would generally report the costs for implementing the programme to the whole population, even the ones who were not included in the experiment. This issue was particularly critical as the proposed CER model (see earlier sections 2.19 as well as 5.4.4) did also include in the calculation the number of beneficiaries of the treatment, which was not always clear. Some studies used schools and some studies used learners as their units of analysis. Also the exact number of beneficiaries was debatable: should only the direct learners be included as the beneficiaries or also the indirect beneficiaries of the systemic interventions, such as the teachers trained, the management development and the communities benefitting from the infrastructure built? The broader social and economic impact of interventions over the years could thus never be precisely estimated.

We have seen in the previous sections how effect size is very susceptible to context. This is even truer for CER, where time and place play a big role. To some extent, the fact that this study was conducted in the same economy (South Africa), made it easier to standardise financial information with regard to currency, inflation and discount rate. But on the other hand the value of educational services and products could still be quite different across urban and rural areas. Programmes in some provinces (such as Limpopo and Northern Cape) were more expensive simply because of the long distances for service providers to travel between schools, compared to more densely populated provinces with better infrastructure like Gauteng.

Some of the interventions discussed in the meta-analysis were pilot projects or academic experiments, which carried a specifically contained cost. This cost, however, would be very different once the pilot would be expanded to a mainstream education policy. The economy of scale would dictate that much of the costs would go down, but at the same time quality (and thus impact) of the intervention would also reduce. Comparing CER of different programmes of different sizes and scope would therefore not be a fair and accurate exercise.

For all the above considerations it was extremely difficult, to compare the cost-effectiveness of the different education interventions included in this meta-analysis in a fair and empirical manner. The exercise would be much easier if the comparison would be done between the interventions of the same organisation or donor, as the costing parameters would at least be similar. Nevertheless, this would still require a high degree of political will by the commissioning agency to allow for the disclosure of the necessary and detailed financial information required for a proper cost-effectiveness analysis.

5.9. Participation and peer reviw

Similar to the process which occurred with the Anglo Platinum case study discussed in chapter 4, the meta-analysis presented in this chapter was also conducted through close collaboration and partnership with the very stakeholders who were the subject of the research. This included regular dialogue with programme managers, government officials, private investors, evaluators and researchers in South Africa's education sector as well as international academics who had conducted similar systematic reviews.

Through a participatory applied research approach (Neuman, 2006), the researcher sought inputs from many of the above stakeholders into the data collection and data analysis process. Once the initial results of the meta-analysis started to emerge these were presented to a number of different fora, for validation, feedback and broader public debate. Between 2015 and 2016 the following presentations were organized to present the findings fo the meta-analysis:

- Academic seminar at the Wits School of Education (September 2015), to discuss preliminary results with South African experts and evaluators, who have been the authors of the primary studies reviewed in the meta-analysis.
- Presentation at the South African Monitoring and Evaluation Association (SAMEA) bi-annual conference (October 2015), during the programmatic strand on education evaluations.
- Academic conversation at the Wits School of Governance (February 2016)
- Seminar at the Wits Business School (June, 2016) to discuss findings with private sector and corporate social investors (CSI) funding education programmes and school development in South Africa.
- Uganda Evaluation Week Conference (Kampala, June, 2016) to present meta-analysis to broader international evaluators, reviewers and policy-makers involved in education reform in Africa.

Although it had not yet occurred at the time of writing this chapter, a policy debate is being planned with the Department of Basic of Education (DBE), Umalusi and the Human Sciences Research Council, to present the results of the meta-analysis with broader policy-makers from national and provincial government education agencies of South Africa.

The above academic and policy seminars were useful fora to confirm the results of the meta-analysis and to receive further inputs and suggestions from peers (evaluators and academics) as well as education practioners on the gaps in the research. Much of the paper presented above has benefitted from the contribution of various South African and international fellow researchers and practioners. Once again the participatory process adopted throughout this research was useful and conducive to the 'democratisation' of knowledge (Stoecker, 1999) being generated and the cross-fertilisation of scientific enquiry with public policy, education programming and development practice.

5.10. Concluding summary

South Africa is still challenged by serious educational backlogs that government and private sector have tried to compensate for through large investments, innovative education models, complex school development programmes, bold education policies and numerous other initiatives implemented by a diversity of institutions. Notwithstanding the financial, technical and political resources invested in the education sector, this study has confirmed that very few interventions have had a substantial impact on learning outcomes in South Africa's public school system (Bloch, 2009; Fleisch, 2008; Kanjee et al., 2013; Taylor et al., 2003). This meta-analysis does not take a formal position with regard to the merits of any particular pedagogical approach, school development programme or learning theory. It rather introduces empirical and systematic methods of evaluation and comparative analysis to the field of education programming.

This meta-analysis is a contribution to evidence-based education research and policy in South Africa. It has captured many of the South African impact studies that international reviews had not previously picked up (Evans & Popova, 2015) and has advanced the work of Mouton et al. (2014) by providing the first meta-analysis of interventions to improve learning outcomes in South Africa's schooling sector. Numerous impact evaluations have been conducted on education programmes in South Africa, with different methods, degrees of quality and scientific rigor. A number of reviews have also been undertaken (Kanjee et al., 2013; Mouton et al., 2014; Schollar & Roberts, 2008) to reflect on the experiences and the learning that have emerged in South Africa's education sector post-1994.

Nonetheless, education policy-makers and investors require standardised impact measures to allow for more robust empirical comparison of the effectiveness of different options available to them.

This chapter has presented the application of a methodological framework for the comparison of intervention effectiveness within the education policy arena of South Africa. Through the use of multiple treatment meta-analysis, the study has analysed the impact of some of the major interventions implemented in public schools over the past 15 years. It has identified some of the programmes that have been most successful in improving learning outcomes and the context and circumstances in which these effects have occurred. While not being conclusive on what are the best education programmes, it has shed some light on when, where, how and why some interventions have shown a significant impact on learner achievement.

What has emerged strongly is that context, methodology and study design have a substantial influence on the effect sizes that are reported in education impact evaluation. Assessment tools, evaluation methods, units of analysis and other methodological choices all appear to influence the type of results that emerge from these impact studies. The focus of the intervention (Slavin, 2013), the timing in which evaluations are conducted (Clemens et al., 2004; Harris, 2008) and potential Hawthorne effects (Schollar, 2013) are all factors that influence impact evaluations. The findings from this South African meta-analysis resonate with similar findings emerging also from a recent American education meta-analysis conducted by Lipsey et al. (2012). It also confirms several of the findings by Evans and Popova (2015), where diverging results of different reviews are dependent on the sample of studies included and the intervention categorisation utilised by the different reviewers. The South African meta-analysis has also shown how results of impact evaluations cannot be taken at face value, as context and study design have a significant influence on the effect sizes reported in the various programme evaluations.

A vivid illustration of the main story emerging out of the South African meta-analysis is the case of the evaluation of the Back to the Basics (B2) workbooks, where the methodological approach undertaken by two different evaluators (Fleisch et al., 2011; Schollar, 2015) led to diametrically opposite results, though they both implemented rigorous RCTs of the same intervention on the same grade level of learners in different provinces of South Africa. Both experiments, though well implemented, had methodological problems and limitations. A third RCT of the B2B workbooks by a different researcher in a new context would definitely be a project worth pursuing in order to reach more definitive conclusions on the impact of guided workbooks in South African primary schools.

”Buig die boompie terwyl hy jonk is”[15]

- Afrikaans popular saying

The strongest predictor of effect size in this meta-analysis has been the schooling phase in which the interventions were implemented and the grade in which the learning assessments were conducted. Programmes implemented at lower levels of the school system report much larger learning impact than programmes implemented in higher grades (i.e., Fleisch & Schoer, 2014 versus Taylor & Watson, 2015). This is consistent also with the general child development literature and some of the international meta-analytical work done in the field of education (Bloom et al., 2008; Hill et al., 2008; Lipsey et al., 2012).

The only exception to the above finding has been the Dinaledi initiative, which has produced remarkable results on science learning outcomes in the FET band (high school), which is an education phase traditionally known to report very small gains in learner achievement. The World Bank (2010) evaluation of Dinaledi, however, suffered from some data limitation and potential selection bias, therefore it might be appropriate to conduct a replication of this study six years later, as the government flagship programme has meanwhile evolved and the Department of Education data has also significantly improved.

Several evaluations of high school interventions in South Africa (Besharati, 2014; Blum et al., 2010; Schollar & Mouton, 2014) have discovered that major learning gains occur when there is an injection of additional human resources and teachers into the schools that help reduce class size and teacher-learner ratios. Numerous studies conducted in primary schools have highlighted the importance of interventions in the area of language (i.e., Learning for Living, READ, RCUP), which produce important gains not only on language outcomes but also on other learning areas (Bohmer, 2014; Fleisch et al., 2015; Schollar, 2005). In South Africa, well-designed and well-delivered learning and teaching resources (LTSMs) appear impactful as well as cost-effective in improving learning outcomes relatively quickly compared to more complex and expensive integrated school development programmes. Some of such examples have been the impact produced by the matric study guides (Taylor & Watson, 2015) and the mathematics workbooks developed by programmes such as the PMRP (Schollar, 2015) and GPLMS (Fleisch & Schoer, 2014). As a caveat, the effectiveness of these materials could also be as a result of the teacher training, lesson planning, coaching and monitoring processes which usually occur as part of the instructional programme delivery model.

[15] Literal translation: “Shape the tree while it is still young” – quoted by Dr. Cas Prinsloo (HSRC) at a seminar in Johannesburg (7 September 2015) commenting on some of the key insights emerging from the meta-analytical study.

This meta-analysis has also shown that experimental evaluations conducted by researcher in carefully controlled environment, such as the testing of innovative teaching approaches or special learning aids (Bohmer, 2014; Hattingh, 2003; Louw et al., 2008; Padayachee et al., 2011) produce high effect sizes, however there is no guarantee that the same impact results are to be expected when the interventions are transferred to the broader policy environment, where beneficiaries often receive partial or incorrect exposure to the treatment. High effect sizes in academic studies can be due also to to publication bias towards studies with positive results and to some degree of Hawthorne effects.

Finally, this chapter has underlined the need for more empirical evidence in South Africa's education sector, and the use of rigorous counterfactual evaluations, such as RCTs and strong quasi-experimental designs, which address selection bias and allow for rigorous analysis of causality. Attention needs to be given to sampling processes to provide more power and confidence in the results of education impact studies (McEwan, 2014; Fleisch et al., 2015). It is hoped that this meta-analysis inspires both public and private agencies to improve the quality of their evaluations, and to utilise the best available scientific methods to systematically plan future programmes, make smarter investments and improve their on-going efforts in the education sector.

Chapter 6: Conclusion

6.1 Evaluation framework emerging from the research

Throughout the decades, development planning and policy has become a more complex social science. International, national and local institutions; public, private, and non-governmental organisations have invested large amounts of resources, technical expertise and endeavoured to implement programmes and policies to improve the social-economic welfare of communities. Development results however often fall short of what was originally planned and expected for, as was the case with the Anglo American education programme in South Africa illustrated in chapter 4.

Evaluation is a very powerful instrument used in development management, that helps answer questions and provide evidence for policy-makers. The research has explored the use of evaluation within the context of South Africa's education sector. Like in the broader development sector (Bamberger et al., 2010; Masud & Yontcheva, 2005; Riddell & Kruse, 1997), this research has highlighted that much of the evaluation practices by public and private institutions still remains weak and lacks the scientific rigor, to produce high-quality empirical evidence required for public policy, improvement of delivery system and effective development programming. Evaluation, especially in South Africa's education sector, has relied heavily on qualitative methods, providing subjective perspectives and is often susceptible to different type of biases (see sections 2.5 and 5.5.3). Before-and-after assessments are very limited, and even fewer evaluations make use of appropriate counterfactual techniques required for credible impact assessment (Fleisch et al., 2015; Mouton et al., 2014,).

The research started by illustrating the development landscape, with the choices, challenge and knowledge gaps that policy-makers currently face. In chapter 2 an in-depth review was conducted of some of the main approaches, methods and techniques used for development evaluation, building on the insights from the literature and experiences from the field, to evolve an evaluation framework that can assist practitioners in the generation of evidence required for effective development programming. The research proposed and piloted a conceptual and methodological framework for the measurement and comparison of different types of interventions to improve education outcomes in South Africa, discussed in detail in chapters 3, 4 and 5.

Mindful of the implications that public policies have on the lives of millions of people, the evaluation framework advocates approaches that encourage empirical rigour in the scientific enquiry around the impact that development interventions have on social-economic conditions of populations and communities. The framework draws on many methods borrowed from health sciences and development economics, which favour a positivist approach to the generation of knowledge. It answers the research questions raised in section 1.3 by proposing the adaptation of techniques like MTMA to allow for the standardisation and comparison of effectiveness measures across a variety of different interventions, agencies and institutions. To address issues around causality, attribution, and validity, the framework suggests the use of theory-based counterfactual evaluation (Mouton et al., 2014; White, 2009), integrating qualitative methods with econometric and quasi-experimental methods, which are easier to undertake in the social sciences, in development and education policy. The framework also integrates elements of participatory applied research, to promote engagement of stakeholders in the collection, analysis and utilisation of evaluation findings and the stimulation of collective learning in the policy context under investigation.

Overall this research has made a contribution to the field of evaluation methods, particularly impact evaluation and systematic meta-analysis, and their use in development management and education policy in South Africa. It has uncovered evidence emerging from South African studies and evaluation reports which was not previously captured by other systematic reviews of education interventions in developing countries (Evans & Popova, 2015), that tend to have a bias towards academic literature generated by Northern scholars, as was shown in chapter 5. It has also advanced previous South African reviews (Mouton et al., 2013) by introducing meta-analysis as an instrument to quantitatively compare effectiveness of different development interventions, and thus assisting policy-makers and donors with decisions on which programmes to favour within a particular development context. On a minor level, the research has also generated some knowledge from South Africa that informs current global development debates; such as the catalytic role that private investments play in supporting education outcomes, and on appropriate monitoring and evaluation systems to assess the contribution of new forms of development partnerships, approaches and interventions in Africa.

This concluding chapter will have a two-fold purpose: it will distil the learning and experiences from piloting the proposed evaluation framework to the two case studies (chapter 4 and chapter 5) in South Africa's education sector; it will also summarise some of the key insights and findings that emerged from the empirical chapters to offer evidence to policy-makers, programme managers and investors active in the transformation of South Africa's schooling system.

6.2 Key findings for South African education policy

The set of methods and instruments set out in the conceptual framework were illustrated within the policy context of South Africa's education sector, laid out first in chapter 3 sectorial overview, and later explored in more depth in the two empirical studies in chapters 4 and chapter 5, where specific school development interventions were evaluated. Chapter 4 assessed the impact of the R 100 million social investments of Anglo Platinum Corporation, aimed at improving mathematics and science learning outcomes in Limpopo and North West schools close to their mining operation sites. Chapter 5 was a comparative meta-analysis of the effectiveness of some of the major interventions aimed at improving learner achievement in South African public schools in the post-Apartheid dispensation; looking also at the factors and context that led to the increase in learning results.

Notwithstanding, the vast volume of education research conducted in South Africa (Chisholm, 2004; Fiske & Ladd, 2004; Motala et al., 2014; Spaull, 2013; Taylor et al., 2013), several scholars (Bloch, 2009; Sayed et al., 2013; Taylor et al., 2003) have stated that there is still little information on the success and failures of programmes implemented in South Africa's schooling system, and on the lessons that can be distilled for public policy. Reliable evidence is scarce as there is little use of rigorous methods of investigation (Motala & Pampallis, 2005) and counterfactual evaluations which address biases and causality (Fleisch & Taylor, 2015); as well as frameworks that allow for the comparison of the effectiveness of different types of interventions aimed at addressing the complex education challenges of the country.

Although the aim of the research was to pilot the methodological framework in a specific development context, the case studies revealed interesting findings for public and private agencies involved in South Africa's education sector. While at the outset the researcher had limited knowledge on the intricate dynamics of the sector, the testing of the evaluation framework unravelled many insights for education policy, which in South Africa and in other developing countries, is central to broader social, economic and political transformation processes. The empirical work undertaken in this research confirmed previous theories, strengthened existing literature, encouraged discussion and learning among stakeholders. It produced important knowledge for South African education policy-makers and investors. Some of the key sectoral findings emerging from the research are summarised as follows:

- The analysis conducted in chapters 3, 4 and 5 has highlighted the substantive role that corporate sector has played in complementing government's efforts in addressing education challenges in South Africa (Besharati, 2015). Notwithstanding the manifold interventions implemented by both public and private institutions, very few programmes have demonstrated a significant and substantial impact in improving learning outcomes in South African schools, that would warrant

serious policy uptake (see chapter 5 meta-analysis).

- In line with other South African literature (Shepherd, 2001; Van der Berg, 2008; Simkins, 2010; Taylor, 2011; Taylor, 2013), the Amplats case study in chapter 4 showed that social-economic factors, such as race, school quintile, school fees, have a bigger influence on learning outcomes than any of the education interventions implemented in the two Northern provinces.
- The Anglo Platinum case study also revealed a surprising psycho-social phenomenon, where schools close to the mining sites, while receiving more attention by corporate investors, suffered at the same time a decrease in mathematics and science pass rates. This was possibly caused by the side effects and social disruptions that mining communities face, as well as from the perverse incentives that encouraged more ill-prepared, ill-equipped and ill-supported learners to enrol (and thus fail) the mathematics and science NSC examinations. Like in other studies (Fleisch, 2006; Prinsloo & Kanjee, 2005), the impact evaluation of the Anglo Platinum programmes showed that there is a strong correlation between an increase in pass rates with a decrease in enrolment in high school exit exams.
- Both chapters 4 and chapter 5 have provided strong examples of how improvements in mathematics, science and other learning areas are highly correlated to the language abilities of learners (Botes & Mji, 2010; Schollar, 2005; Simkins, 2010; Taylor & Prinsloo, 2005). This may affect critical decisions around the language policy in South Africa's complex multi-lingual schooling system. This goes to further corroborate other South African research on the topic (Cummins, 2009; Kaphesi, 2001; Myburgh et al., 2004; Taylor et al., 2013)
- As previously seen also in the North American literature (Bloom et al., 2008; Hill et al., 2008; Lipsey et al., 2012), the meta-analysis undertaken of South Africa's education programmes in chapter 5, revealed that impact of interventions is likely to appear larger if implemented in lower grades, where learning gains of younger children are generally expected to be higher than those of older students because of normal trends in cognitive capabilities.
- Both the Anglo Platinum case study in chapter 4 and the meta-analysis in chapter 5 have shed further evidence on the highly debated government flagship programme of Dinaledi. This programme appears to have exceptionally high impact on mathematics and science results compared to other interventions implemented at secondary school level. The previous World Bank evaluations of the Dinaledi (Blum et al., 2010) suffered from potential bias and data limitations, therefore it is worth conducting a follow-up impact evaluation of the programme, as this has also evolved since its original design.
- Quantity and quality of teachers is a critical factor for the functioning of South Africa's education system (Chisholm, 2009; Simkins, 2010; Spaull, 2011; Taylor, 2011; Wolhuter, 2006).

Interventions with educators can thus provide better, long-term and sustainable results for the overall schooling system. Regressions in chapter 4 show a negative correlation between increased teacher-learner ratio and a decrease in learning results of schools, thus confirming the importance of class size for quality teaching and learning to occur. Some of the successes of programmes such as Dinaledi, GPLMS and Plus Time, analysed in chapter 5, could also be attributed to the provision of additional educators, training, coaching and teaching support material. Quality of South African teachers is closely related to the quality of the pre-service and in-service educator development programmes (Ono & Ferreira, 2010; Parker, 2002; Wolhuter, 2006). Further empirical research is required around interventions to improve teaching quality and greater attention needs to be given to this topic by policy-makers and investors in South Africa.

- Linked to the quality and performance of teachers, is the central role of school leaders, managers and governing bodies, which play a critical role in promoting local accountability, counterbalancing pressures of unions, and ensuring school functionality and effectiveness. The findings from the field work conducted in Limpopo and North West in chapter 4, confirmed much of the literature on this subject (Christie, Butler, & Potterton, 2007; NPC, 2011; Taylor et al., 2003; Taylor, Gamble, Spies & Garisch, 2013).

6.3 Reflections and learning from the application the evaluation framework

Aside from the important findings for education policy emerging from the empirical work conducted in this research, this conclusion chapter responds to the initial research questions in section 1.3 and discusses the experience and lessons learned in the application of the methodological framework in South Africa's education sector. This section will summarise the main features of the evaluation framework, discuss its merits, strength, weaknesses, and limitations, its relationship to the literature, and the contribution it makes to knowledge in field of impact evaluation and systematic reviews. The experiences and lessons gathered in the initial pilot studies, allows for modification and refinement of the methodological framework in future research and other follow-up evaluations in other development sectors and country contexts.

6.3.1 Measuring impact

Throughout this research, the methodological weaknesses, the lack of empirical rigour and limitations of current evaluation practice prevailing in the international development industry (Bamberger et al., 2010; Masud & Yontcheva, 2005; Riddell & Kruse, 1997), and likewise in South Africa's education sector (see chapters 4 and 5), have been repeatedly underscored. The evaluation of the Anglo American education programme presented in chapter 4 has illustrated a typical local example of the 'micro-macro paradox' discussed in previous literature by development economists such as Mosley (1986) and Picciotto (2006). The South African case study showed that a rigorously performed impact assessment conducted on the results of an education intervention on outcomes at provincial and school level reached diametrically opposite results from the results that emerged from micro-level evaluations of programme beneficiaries done by the service providers involved in the implementation of the Anglo American programmes. Seemingly successful results of small interventions did not translate into wider improvements at the system level. This was due to potential funding and selection bias, and to the weak scientific rigor of many agency-level evaluations of South African education programmes (Fleisch et al., 2014; Schollar, 2015).

The meta-analysis in chapter 5 also highlighted that education evaluations in South Africa still remain relatively weak. The majority of evaluations tend to rely heavily on qualitative (and thus more subjective) approaches to analysis, and very few undertake rigorous counterfactual evaluations required to produce credible evidence for development planning and policy. One of the biggest factors contributing to the micro-macro paradox is that aside from the intervention itself there are always many other confounding factors and forces that influence the development outcome of interest. The econometric, experimental and qualitative analysis in chapter 4, for example, discovered important psyo-social and environmental factors present in the mining communities that had larger influence on school results than the well-intended education programmes of Anglo American and other development agencies.

Addressing the problem of causality, attribution and isolating the net effects of a programme from the effects of all other interventions and factors in the environment (research sub-question 1a) is a problem which has been discussed at length in evaluation literature (Gertler et al., 2011; Johnson & Lamdany, 2005; Mouton et al., 2013; White, 2005). This is normally addressed through the construction of a counterfactual that can indicate the condition of the beneficiaries with and without the programme. Experimental approaches, like RCTs, have long been considered the 'gold standard' for impact evaluation promoted by development economists such as Baker (2000), Kremer et al. (2013), Duflo, Glennerster, and Banerjee (2009).

As discussed in chapter 2, randomised experiments come however with many constraints with regard to cost, time, ethics, and often they are unpractical to be undertaken in development policy settings, where evaluations are done ex-post and randomised assignment of participant is no longer possible. Chapter 5 meta-analysis has shown that the majority of impact evaluations in South Africa's education sector utilise quasi-experimental methods, such as matching techniques, natural experiments, cut-off point designs, double difference methods, to conduct counterfactual evaluation. The impact evaluation of Anglo American education programme in chapter 4, was also made possible through combining DID method with PSM and RDD. Similar to the natural experiment previously conducted by Gustafsson and Taylor (2013) with the change of administrative boundaries in South Africa, the evaluation of the impact of mining operations on neighbouring schools, conducted in chapter 4, allowed for the experimentation of geo-spatial approaches to cut-off point design. The quasi-experimental methods used were unobtrusive on the population, much cheaper, and more practical for the retrospective evaluations, which are often conducted in dynamic development policy contexts, as was the case of this research.

The impact evaluation of the Anglo American programme also confirmed earlier literature (Dehejia, 2013; Hombrados & Waddington, 2012; Shadish & Cook, 2009) that explains how well implemented quasi-experimental designs, that address selection and other biases through careful statistical controls, produce powerful results that are just as robust as the ones reached through RCTs, and thus offer a useful and cheaper alternative to policy-makers, investors and development practitioners - thus responding to sub-question 1b of this research.

One of the strengths of the evaluation framework utilised in the research was the mixing and combining of different quantitative and qualitative methods to allow for triangulation (Bambergeret al., 2010) and the production of superior evidence (Bryman, 2007). To reduce potential bias and strengthen both external and internal validity, different sets of methods were combined in the impact evaluation work. Chapter 4 provided a good example of how the combination of sophisticated quantitative methods and qualitative field-work assisted in revealing, explaining and confirming the same findings through different angles and further strengthening the evidence produced in the case study. In the Anglo Platinum impact study, it was the interviews with the school principals that helped explain the unusual trends emerging from the statistical analysis of math and science pass rates in the communities around the mining sites. The integration of qualitative methods in the evaluation framework was also very important in order to avoid a 'black box' approach (Bickman, 2000; Chen & Rossi, 1987; Lipsey, 1993). Using mixed methods provided depth, context and further insight to the evaluation, providing explanations about why and how certain phenomena would occur.

In response to sub question 1c[16], what emerged clearly in chapter 5 meta-analysis was that context and study design matter a great deal, and that such factors sometimes have a bigger influence on effect size than the intervention itself. Methodological factors (such as units of analysis, assessment instruments, evaluation methods) can therefore mask the true impact of programmes (Lipsey et al., 2012), and be deceiving if not properly controlled for. As illustrated in previous works by Clemens et al. (2004), Harris (2009), and Schollar (2015), the above research confirmed that the timing of when impact evaluations are conducted makes a substantial difference in the magnitude of the impact that ends up emerging from the evaluations. When conducting impact assessments, evaluators need to be clear on when the impact of a specific type of development intervention is expected to mature, and that results may differ in the short (immediately after the project), medium and longer term (5 years later). This confirms previous studies (Andrabi et al., 2011, Evans, Kremer & Ngatia, 2014) showing that impact results of interventions are normally not sustained over time.

The research also showed how contextual factors can taint substantially the results of impact evaluations. The meta-analysis of South African education programmes confirmed international literature (Bloom et al., 2008; Hill et al., 2008; Lipsey et al., 2012) that states that interventions undertaken with younger children at lower grades, are likely to produce bigger impact on learning outcomes than interventions done with older students at the end of the schooling system. Chapter 5 also illustrated the interesting case of two impact studies (Fleisch et al., 2011; Schollar, 2015) of the same intervention (B2B Workbooks), implemented in the same manner, but in different South African provinces and with slightly different experimental designs that produced diametrically opposite results from one another.

The meta-analysis has also showed that small academic studies conducted in carefully controlled experimental environments (Hattingh, 2003; Louw et al., 2008; Padayachee et al., 2011) are likely to produce much bigger effects than public sector evaluations of large system-wide programmes implemented by national and provincial governments. Thus scale of programmes influences also the effect size expected to emerge from the evaluative studies. The econometric analysis in the Amplats case study in chapter 4 also corroborated other South African education impact studies (Blum et al., 2010; Gustaffson & Taylor, 2013; Simpkins, 2010; Van der Berg, 2008) that show that geography (i.e., the former homeland areas, provincial administrative lines) have also some influence on learning outcomes of South African public schools.

[16] What are the exogenous and endogeneous factors that influence the magnitude of the impact caused by a development intervention?

6.3.2 Comparing effectiveness

Similar to the impact evaluation work discussed above, also in the systematic review conducted as part of this research, a fine balance was kept between quantitative and qualitative approaches. Most of the South African education reviews, including the latest Mouton et al. (2014) review, can be characterised as narrative reviews (Davies, 2015). While these can be useful approaches to explore the mechanisms of change of interventions and systems, they are also subject to biases and to authors' prior belief system (Evans & Popova, 2015). They are also limiting for investors and policy-makers who require clear answers on the effectiveness of programmes in relationship to other alternatives.

Comparing results from quantitative evaluations, on the other hand, is problematic as different studies utilise different instruments, scales and systems of measurement (Donaldson, 2005). This was illustrated also in the systematic review in chapter 5. To allow for comparison between evaluations of different programmes and policies, standardising results was thus a critical step in the process. Using meta-analytical measures like SMD, Cohen's *d* and Hedge's *g* proved to be an effective way to address research sub-question 2a, thus providing a useful approach for the comparative analysis of the effectiveness of school interventions, as was shown in chapter 5. More advanced approaches such as MTC, MTMA, and NMA, introduced by Caldwell (2005), Lu and Ades (2006), and Salanti, (2007) were found more appropriate approaches for the structured comparison of different types of interventions, which was the case of the comparative study conducted in this research.

The MTC in chapter 5 essentially emulated a large multiple treatment experiment, by utilising a common control group across studies, consisting of the general South African public school system affected by the same political, economic and social forces operating in the country (business as usual). Comparisons were made through observational means, by assessing one intervention in relationship to another intervention against the null (control) situation. This model was based on the assumption that the comparator used to assess the effectiveness of the various interventions was somewhat similar and consistent across studies. The forest plots, commonly used to represent the results of meta-analyses, provided also a powerful and user-friendly tool for policy-makers and investors to easily and visually interpret the results of the statistical work and make eyeball comparisons of the effectiveness of the various programmes - thus fulfilling research sub-question 2b.

Chapter 5 also highlighted the key assumption of meta-analysis with regard to comparing 'apples and apples' - that interventions and evaluations need to be conceptually, methodologically and statistically similar (Duvendack et al., 2012). It was therefore crucial to clearly delineate the inclusion criteria for the meta-analysis and apply it consistently to all the studies being reviewed. When comparing, synthesising and pooling together the results of the different studies it became evident that

methodological quality of each evaluation had to also be carefully assessed to ensure that few bad studies did not skew the overall results of the meta-analysis (Hombrados & Waddington, 2012).

Evans and Popova (2015) have shown that heterogeneity of results is to be found even among systematic reviews, which are susceptible to the way interventions are categorized, to the sample of studies included in the review, as well as other contextual drivers. The systematic review conducted in chapter 5 combined narrative (qualitative) approaches with meta-analytical (statistical) approaches, and uncovered 27 South African impact studies which had not been captured by any of the previous international reviews on education in developing countries (Evans & Popova, 2015), thus signaling an ongoing publication bias towards Northern academic literature.

6.3.3 Participation, peer-review, policy and practice

As the research unfolded an evaluation framework gradually emerged which could assist development policy-makers to gather empirical evidence on what works and what does not, to asses how much impact interventions produce, and decide which are the best programmatic option to undertake when endeavouring to achieve a specific development objective. While being grounded in scientific standards and approaches, the framework provided also a practical tool for development practitioners operating in an evolving policy environment, faced with considerations of time, resources, ethics and context – as called for in the overarching research questions in Section 1.3.

Another key features of the methodological framework piloted in this research was the integration of participatory approaches (Burawoy, 1998; Schafft & Greenwood, 2003) and elements of applied research (Neuman, 2006). The research engaged a range of stakeholders – evaluators, beneficiaries, donors, implementing agencies - in the analytical endeavour and encouraged wide debate on the findings from the two empirical studies. This was seen in the Anglo Platinum case study, where provincial and national government authorities, private investors, service providers, CSOs, academics, experts, media and local communities, contributed to the enquiry through the provision of data, but also through the analysis and dissemination of the findings. Similarly, in the meta-analysis in chapter 5 various evaluators and scholars provided crucial data, expert inputs, and later assisted to confirm and interpret the results stemming from the systematic review of South Africa's education interventions, through an ongoing process of peer review.

Once evidence was produced through rigorous research procedures, results were synthesised, simplified and presented to a diverse set of partners for comments, criticism, validation and further inputs. This allowed for buy-in from participants, for more transparency and democratisation of the research process, for the dissemination of findings into the public domain, collective learning and the

cross-fertilisation between academia and development practice. The knowledge generated in the study was thus jointly owned by the researcher, the practitioners and stakeholders, and transferred more swiftly into insights and learning for South African education policy.

The conceptual framework presented in this thesis has tried to bridge the world of evaluation research (Neuman, 2006) and evaluation practice. In real life, an evaluator would rarely have the luxury (time, skills, and resources) to go through all the steps presented in this research; nonetheless, the framework encourages evaluators to take a scientific posture, and use in a systematic way the enquiry methods presented in chapter 2 (and illustrated further in the examples in chapter 4 and 5), to respond to ongoing questions and decisions that policy-makers and stakeholders need to routinely make about social programmes they are engaged with.

6.3.4 Challenges and limitations

The evaluation framework developed and used throughout this research offers powerful instruments to empirically measure and compare the effectiveness of development interventions. Its relies on many quantitative approaches such as meta-analysis, quasi-experimental designs, and the statistical operations that accompany these methods, like effect size calculation, PSM, multivariate regressions, ANOVA and significance testing.

All these methods, however, are dependent on large volumes of data that combine programmatic, developmental and contextual statistics into large datasets that need to be carefully organised, cleaned and managed. Data quality and data availability are thus critical for the proper implementation of the evaluation framework. Information about the programme (its features, its implementation, costs and beneficiaries), the development outcome of interest (with its performance indicators), as well as other social-economic and environmental variables from the context under analysis, all need to be collected at micro-level and with regular frequency in order to effectively utilise some of the methods proposed above.

Large datasets were also critical for the sampling process in order to generate enough statistical power for the impact evaluations. In the quasi-experimental designs undertaken on various aspects of the Anglo Platinum interventions, the larger was the sample size of both treatment and control groups, the stronger was also the internal and external validity of the results emerging from the evaluations. The meta-analysis in chapter 5, also highlighted how sample size had a large influence on the precision of the effect size calculated for the various interventions. The larger the sample size was, the smaller the standard error and shorter range of confidence interval in the impact results reported in the meta-analysis. When sample sizes were too small the evaluations could not produce statistically significant results that

would provide reliable evidence on the impact of the interventions, as was the case of the Radical M&S winter school analysed also in chapter 4.

The piloting of the methodological framework in South Africa's education sector, was made possible in great part to the data available through the different South African institutions who have been collecting micro-level education statistics for many years (i.e. NSC exam results from Umalusi, school characteristics from EMIS databases of the DBE, socio-economic features of communities from StatsSA). Although still imperfect, the availability of data as well as the vast pool of existing research done in the sector, made it possible to implement the proposed evaluation framework in South Africa's education arena. This however may not always be the case. Other African and developing country contexts may face much more severe data constraints and weaker institutions than in South Africa's stable middle-income economy. Therefore, secondary data required for the type of quantitative analysis discussed above might not always be possible to the same degree because of more limited or non-existent data for the specific sector or country under analysis.

In development policy, cost implications have become a very important factor in decision-making (research sub-question 2b), therefore the research endeavoured to the integrate also CEA into the evaluation framework in order to assist in gathering evidence on value-for-money and best interventions options to undertake in a specific context. Development programmes that cost less, reach more participants, and have a bigger effect will normally be favoured against the more expensive, less impactful and smaller in scope. A cost-effectiveness index would thus allow policy-makers to compare more easily interventions and identify the ones that are more suitable for replication and capitalisation.

Implementing the cost-effectiveness component of the framework, nevertheless, proved to be much more difficult than originally planned. Although calculating the adjusted CER appeared misleadingly simple and straightforward at first, retrieving accurate and complete financial information from public or private institutions was an arduous, sensitive and painstaking exercise. Even when budgetary information was made available, it often was not reported in a consistent manner across the various programmes and institutions.

Doing any form of cost-analysis would require a thorough breakdown and scrutiny of all the costs, direct and in-kind, to the implementing organisations, to the partners as well as the users of the programme – procedures which would be prohibitive to do with the average evaluation resources, time-frame and expertise available to development organisations. Even with strong political commitment to transparency and availability of technical expertise, the best cost information that could be gathered would still be only estimates, and this would detract from the scientific validity of the exercise.

The above challenges were illustrated in the comparative meta-analysis of South Africa's education programmes in chapter 5. What transpired was that it is easier to conduct a CEA when there is strong institutional backing from the same institution (i.e. a specific donor or government agency), who conducts multiple initiatives within the same context, thus making it easier to compare costs as they are at least reported in a similar manner. Although not impossible, integration of CEA in the evaluation process is extremely difficult, and would certainly require more attention, time, resources and work in subsequent research.

6.3.5 Future application to other contexts and policy sectors

What emerged out of the above experience is that for effective utilisation of the proposed conceptual framework certain preconditions are required in the sector and in the country in which the analysis takes place. Firstly, the area of exploration needs to be politically significant and of high interest to policy-makers and investors, who ultimately need to provide the resources and political support to allow for the complex and often sensitive data gathering to occur. As the model provides a comparative framework, it is more meaningful if there is a multiplicity of activities and institutions in the same terrain addressing the same objectives, which can eventually be compared against one another. The development outcome needs to be clearly specified, and the theory of change unpacked to explain how the interventions affect change in the outcome in the short or medium term, with at least some level of expected attribution.

As the methodological framework relies heavily on quantitative approaches such as quasi-experimental designs, CEA, econometric and meta-analytical methods, it is crucial that quantitative indicators are available to measure progress or regress on the outcome of interest. Reliable social-economic statistics need to be available regularly and at the lowest geographic level as possible. To allow for the use of some of the impact evaluation techniques and to justify the costs of some of the complex evaluation procedures discussed in chapter 2, programmes evaluated need to be sizeable and affecting a large number of participants.

The empirical work undertaken in South Africa's education sector, offered a real-life policy setting with many of the above pre-conditions, thus providing a good testing ground for the conceptual framework proposed in this research. Education and skills development features as a very high priority for both national and provincial governments, as well as for private investors in South Africa. Numerous initiatives, programmes and policies have been implemented in South Africa by diverse institutions (NGOs, government agencies, private foundations, foreign donors) to address the common development challenge of improving learner performance in public schools.

Addressing attribution, however, is particularly challenging in education programming, as there are many factors, internal and external to the school system, which affect the outcome of improved learning results (see also chapter 3). The use of counterfactual evaluation methods therefore played a critical role in the analyses conducted in chapter 4 and 5. Fortunately, South Africa's education sector presented a policy space very rich in data that allowed for the use of several learner test data (NSC exams, ANA results, HSRC and JET assessments) combined with extensive administrative and demographic data available down to school level thanks to the databases of the Department of Education and Statistics South Africa.

The experiment of implementing the conceptual and methodological framework in South Africa's education sector was overall successful, though it was clear that data availability played an important role in the process and clear constraints still exist with the cost-effectiveness component of the model. The pilot studies in South Africa showed that the framework can be utilised to empirically measure the impact of a variety of different types of interventions, agencies and institutions and to compare effectiveness across a variety of different studies and evaluations through a meta-analytical process.

The lessons gained throughout this research can be now extracted to modify, further refine and adapt the framework to new geographic contexts and policy sectors. Following the initial pilots in the education field, other potential sectors to test the framework in the future could include health, agriculture, infrastructure, and employment - which are all important areas in South Africa's as well as Africa's continental development processes (NEPAD, 2002). It would also be appropriate to explore further applications of the framework to other geographic, political and social-economic contexts, such as low-income countries and post-conflict/fragile states, where capacities are weaker and data less available and reliable, than in the pilots conducted in this research.

The framework remains an on-going 'work-in-progress' and a continuous learning journey. It will continue to evolve and improve as further conceptual inputs emerge from the academic literature, from the insights of other experts, as well as from the experiences gathered from further field experimentation. The framework will continue to be tested and applied to different development settings, populations and sectors, in order to see if and how it works under different conditions. Insights, experiences and learning gained from the application of the evaluation framework will be used to constantly improve and refine the models, approaches and methods into more powerful and useful tools for development policy-makers and programme managers. Plenty of avenues exist ahead to expand, deepen and develop further the methodological framework as a subject of post-doctoral research endeavours.

Reference list

ActionAid (2005). *Real aid: An agenda for making aid work*. London: ActionAid International.

Adams, W. M. (2003). *Green development: environment and sustainability in the Third World*. New York: Routledge

Agency for Healthcare Research and Quality (AHRQ) (2012). *Comparative effectiveness portfolio*. Retrieved on 23/07/2016 from http://www.ahrq.gov/cpi/portfolios/comparative-effectiveness/index.html

Addison, T., Mavrotas, G., & McGillivray, M. (2005). *Aid, debt relief and new sources of finance for meeting the Millennium Development Goals* (WIDER Research Paper No. 2005/09). Helsinki: UNU-WIDER.

Alidou, H., Boly, A., Brock-Utne, B., Diallo, Y. S., Heugh, K., & Wolff, H. E. (2006). *Optimizing learning and education in Africa–the language factor*. Paris*: ADEA*.

Africa Mining IQ (2016). *Platinum mining in Africa*. Retrieved from http://www.projectsiq.co.za/platinum-mining-africa.htm

Ahmed, R., & Sayed, Y. (2009). Promoting access and enhancing education opportunities? The case of 'no-fees school' in South Africa. *Compare: A Journal of Comparative and International Education*, *39*(2), 203-218 (Special Issue: Non-state provision of education: Evidence from Africa and Asia)

Alesina, A., & Dollar, D. (2000). Who gives foreign aid to whom and why? *Journal of Economic Growth*, 5(1), 33-63.

Alkire, S. (2002). *Valuing freedoms: Sen's capability approach and poverty reduction*. Oxford: Oxford University Press.

Andrabi, T., Das, J., Khwaja, A. I., & Zajonc, T. (2011). Do value-added estimates add value? Accounting for learning dynamics. *American Economic Journal: Applied Economics*, *3*(3), 29-54.

Anglo American (2012). *A balanced portfolio with the future in mind* (Annual Report 2012). Johannesburg: Anglo American.

Arbab, F. (2000). Promoting a discourse on science, religion, and development. In S. Harper (Ed.), *The lab, the temple and the market: Reflections at the intersection of science, religion, and development* (pp. 149-237). Ottawa: International Development Research Centre (IDRC)

Arndt, C., Jones, S., & Tarp, F. (2006). *Aid and development: the Mozambican case*. Copenhagen: Department of Economics, University of Copenhagen.

Ashman, S. (2013) *Political economy of platinum*, paper presented at conference on The Meaning of the Rise of the South African Platinum Mining Industry and the Nature of the Post-Apartheid Order, Johannesburg: University of Witwatersrand

Asmal, K. (2002). *Minister of Education's quarterly report to the President: fifth report on the provinces to the President from the Minister of Education, Professor Kader Asmal, MP.* Pretoria: Department of Education.

Astbury, B. & Leeuw, F.L. (2010). Unpacking black boxes: mechanisms and theory building in evaluation. *American Journal of Evaluation, 31*(3), 363-381.

AUSAID (2008). *Annual Review of Development Effectiveness 2008: Scaling up in a deteriorating global environment*, Canberra: AUSAID

Badenhorst, C. (2007). *Research writing: breaking the barriers.* Pretoria: Van Schaik.

Baird, S., Ferreira, F. H., Özler, B., & Woolcock, M. (2013). Relative effectiveness of conditional and unconditional cash transfers for schooling outcomes in developing countries: a systematic review. *Campbell Systematic Reviews*, *9*(8).

Baker, J. L. (2000). *Evaluating the impact of development projects on poverty: a handbook for practitioners* (Directions in Development Series). Washington, DC: World Bank.

Bamberger, M. (2009). Strengthening the evaluation of program effectiveness through reconstructing baseline data. *Journal of Development Effectiveness 1*(1), 37-59.

Bamberger, M., Rao, V., & Woolcock, M. (2010). Using mixed methods in monitoring and evaluation: experiences from international development evaluation. In A. Tashakkori, & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social and behavioral research* (2nd ed., pp. 613-641). Thousand Oaks, CA: Sage Publications.

Banerjee, A. V. (2007). *Making aid work*. Cambridge, MA: MIT Press.

Banerjee, A., Duflo, E., Glennerster, R., & Kinnan, C. (2015). The miracle of microfinance? Evidence from a randomized evaluation. *American Economic Journal: Applied Economics*, *7*(1), 22-53.

Becker, B. J., & Wu, M. J. (2007). The synthesis of regression slopes in meta-analysis. *Statistical Science*, *22*(3), 414-429.

Bello, O., & Manrique, M. (2011). Africa rising?. In R. Youngs (Ed.), *Challenges for European foreign policy in 2012: what kind of geo-economic Europe* (pp.103-111). Madrid: FRIDE.

Belli, P., Anderson, J., Barnum, H., Dixon, J., & Tan, J.-P. (1998). *The World Bank handbook on economic analysis of investment operations*. Washington,D.C.: World Bank.

Benkenstein. A. (2013). *The platinum sector peace and stability accord: solution or reprieve?* Retrieved on 26/02/2014 from http://www.saiia.org.za/opinion-analysis/the-platinum-sector-peace-and-stability-accord-solution-or-reprieve

Bennell, P. (2003, December). *The impact of the AIDS epidemic on schooling in sub-Saharan Africa*. Background paper for the Biennial Meeting of the Association for the Development of Education in Africa in Grand Baie. Mauritius, December 2003.

Besharati, N. (2012). *Raising Mozambique: development through coal* (SAIIA Policy Briefing No. 56). Johannesburg: SAIIA.

Besharati, N. (2013a). *A year after Busan: where is the Global Partnership going?* (Occasional Paper No.136 Economic Diplomacy Programme). Johannesburg: South African Institute of International Affairs (SAIIA).

Besharati, N. (2013b). *SADPA: strategic aid or development packages for Africa?* (Research Report No.12 Economic Diplomacy Programme). Johannesburg: South African Institute of International Affairs (SAIIA).

Besharati, N. (2013c). *Common goals differential commitments: the role emerging economies in global development.* Discussion Paper 26/2013. Bonn: German Development Institute

Besharati, N. (2014). *Platinum & passes: the impact of mining investments on education outcomes in South Africa* (Research Report No. 16). Johannesburg: SAIIA & Wits School of Governance.

Besharati, N. (2015). *Private sector investment in South Africa's education sector: nature, extent and incentives* (Research Brief, 2015). Johannesburg: Jet Education Services.

Bertrand, M., Duflo, E., & Mullainathan, S. (2001). *How much should we trust difference-in-differences estimates?* (Mimeograph). Cambridge: MIT.

Bhorat, H., & van der Westhuizen, C. (2008). *Economic growth, poverty and inequality in South Africa: the first decade of democracy*. Cape Town: University of Cape Town (Development Policy Research Unit)

Bickman, L. (1987). The functions of program theory. *New Directions for Program Evaluation*, *1987*(33), 5-18.

Bickman, L. (2000). Summing up program theory. *New Directions for Evaluation*, 2000, 103–112.

Blanchard, O. J. (1987). Neoclassical synthesis. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *The new Palgrave: a dictionary of economics* (pp. 634-637). London: Macmillan Press Ltd.

Bloch, G. (2009). *The toxic mix: what's wrong with South Africa's schools and how to fix it.* Cape Town: Tafelberg.

Bloom, H., Hill, C., Black, AR., & Lipsey, M. (2008), Performance trajectories and performance gaps as achievement effect-size benchmarks for education interventions. *Journal of Research on Educational Effectiveness*, *1*, 289-328.

Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis*, 29(1), 30-59.

Blum, J., Krishnan, N., & Legovini, A. (2010). *Expanding opportunities for South African youth throuh math & science: the impact of the Dinaledi program*. Washington DC: World Bank Development Impact Evaluation Initiative (DIME).

Bogdan, R., & Biklen, S. (2007). *Qualitative research for education*. Boston: Allyn & Bacon.

Bohmer, B. (2014). *Testing Numeric: Evidence from a randomized controlled trial of a computer based mathematics intervention in Cape Town High Schools*. Unpublished Masters Thesis. University of Cape Town

Bold, T., Kimenyi, M., Mwabu, G., Ng'ang'a, A., & Sandefur, J. (2013). *Scaling up what works: experimental evidence on external validity in Kenyan education*. Working Paper, (321). Washington DC: Center for Global Development

Bonell, C. P., Hargreaves, J., Cousens, S., Ross, D., Hayes, R., Petticrew, M., & Kirkwood, B. R. (2011). Alternatives to randomisation in the evaluation of public health interventions: design challenges and solutions. *Journal of Epidemiology and Community Health*, *65*(7), 582-587.

Boone, P. (1994). *The impact of foreign aid on savings and growth*. London School of Economics and Political Science, Centre for Economic Performance.

Boone, P. (1996). Politics and the effectiveness of foreign aid. *European Economic Review*, *40*(2), 289–329.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex: John Wiley & Sons.

Botes, H., & Mji, A. (2010). Language diversity in the mathematics classroom: does a learner companion make a difference? *South African Journal of Education*, *30*(1), 123–138.

Bouffard, S., & Little, P. M. (2004). Detangling data collection: methods for gathering data. *Harvard Family Research Project*, *5*, 1-6.

Bovens, M., Goodin, R. E., & Schillemans, T. (Eds.). (2014). *The Oxford handbook of public accountability*. OUP Oxford.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012). *Estimating the effect of leaders on public sector productivity: The case of school principals* (No. w17803). University of Stellenbosch: National Bureau of Economic Research.

Bray, M., & Lykins, C. (2012). *Shadow education private supplementary tutoring and its implications for policy makers in Asia*. Manila: Asian Development Bank

Bray, M., (2013), Shadow Education: comparative perspectives on the expansion and implications of private supplementary tutoring. *Procedia – Social and Behavioral Sciences*. (77) 412-420

Bray, M. (2013). Benefits and tensions of shadow education: Comparative perspectives on the roles and impact of private supplementary tutoring in the lives of Hong Kong students. *Journal of International and Comparative Education*, *2*(1), 18-30.

Brumm, H.J. (2003). Aid, policies, and growth: Bauer was right. *Cato Journal, 23* (2), 167-174.

Bryman, A. (1984). The debate about quantitative and qualitative research: a question of method or epistemology? *British journal of Sociology*, 75-92.

Bryman, A. (2007). Barriers to integrating quantitative and qualitative research. *Journal of Mixed Methods Research*, 1(1), 8.

Brunton, K. (2006). Education for sustainable development: principles for curriculum development in business subject areas. *Investigations in university teaching and learning*, *3*(2), 36-46.

Buchanan, J. & Tullock, G. (1999). *The calculus of consent: logical foundations of constitutional democracy (*orig. pub. 1962.). Indianapolis, IN: Liberty Fund, Inc.

Buchmann, C (2002). Getting ahead in Kenya: social capiral, shadow education and achievement. In B. Fuller & E. Hannum (Eds), *Schooling and social capital in diverse cultutes*. (pp. 133-159) Amsterdam: JAI press

Buckland, P., & Fielden, J. (1994). *Public expenditure on education in South Africa, 1987/8 to 1991/2: An analysis of the data.* Johannesburg: Centre for Education Policy Development (CEPD).

Buhlungu, S., Daniel, J., Southall, R., & Lutchman, J. (eds.) (2007). *State of the nation: South Africa 2007*. Pretoria: Human Sciences Research Council.

Burawoy, M. (1998). The extended case method. *Sociological Theory*, 16(1), 4-33.

Burch, P. (2009). *Hidden markets: the new education privatization*. New York: Routledge

Burnside, C., & Dollar, D. (2000). Aid, growth, the incentive regime, and poverty reduction. *The World Bank: Structure and Policies*, 3, 210.

Bush, T., & Heystek, J. (2006). School leadership and management in South Africa: Principals' perceptions. *International Studies in Educational Administration*, *34*(3), 63-76.

Bush, T. (2007). Educational leadership and management: Theory, policy and practice. *South African Journal of Education*, *27*(3), 391-406.

Bush, T., Duku, N., Kiggundu, E., Kola, S., Msila, V., & Moorosi, P. (2009). The Zenex ACE: school leadership research: final report. *Pretoria, South Africa: Department of Education*.

Bush, T., Joubert, R., Kiggundu, E., & van Rooyen, J. (2010). Managing teaching and learning in South African schools. *International Journal of Educational Development*. 30 (2), 162-168

Bucher, H.C., Guyatt, G.H., Griffith, L.E., & Walter, S.D (1997). The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *Journal of Clinical Epidemiology 50*, 683–691

Byrnes, R. M. (1997) (Ed.). (3rd ed.). *South Africa: A country study* . Washington, D.C.: Federal Research Division, Library of Congress.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for implementation of propensity score matching. *Journal of Economic Surveys*, *22*(1), 31-72.

Calderisi, R. (2006). *The trouble with Africa: why foreign aid isn't working?* Houndmills: Palgrave Macmillan.

Caldwell, D.M. (2005). Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *British Medical Journal*, 7521, 897.

Callinicos, L. (1993). *A place in the city: the rand on the eve of apartheid.* (Vol3). Raven Press.

Campbell Collaboration (2015). *Campbell Collaboration systematic reviews: policies and guidelines* (Monograph Series, Supplement 1). Oslo: Campbell Collaboration.

Campbell, D. T., Stanley, J. C., & Gage, N. L. (1963*). Handbook of research on teaching: experimental and quasi-experimental designs for research on teaching*. Boston: Houghton Mifflin.

Campbell, D.T., & Stanley, S.C. (1966). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

Campbell, D. T. (1991). Methods for the experimenting society. *American Journal of Evaluation*, *12*(3), 223-260.

Capps, G. (2013, October). *Contrasting configurations of platinum capital and tribal authority: Bafokeng, Bengwenyama and Langa Mapela*, paper presented at the conference on The Meaning of the Rise of the South African Platinum Mining Industry and the Nature of the Post-Apartheid Order, Johannesburg: University of Witwatersrand.

Carnevale, A. P., Rose, S. J., & Cheah, B. (2011). *The college payoff: education, occupations, lifetime earnings*. Washington, DC: Georgetown University Center on Education and the Workforce.

Carnoy, M., Chisholm, L., Addy, N., Arends, F., Baloyi, H., Irving, M., & Sorto, A. (2011). *The process of learning in South Africa: The quality of mathematics teaching in North West Province*. Technical report. Pretoria, SA / Stanford, CA: Human Sciences Research Council (HSRC)/School of Education, Stanford University.

Carnoy, M., Chisholm, L., & Chilisa, B. (2012). *The low achievement trap*. Pretoria: Human Sciences Research Council (HSRC) Press.

Carvalho, S., & White, H. (2004). Theory-based evaluation: the case of social funds. *American Journal of Evaluation*, *25*(2),141–160.

Case, A., & Deaton, A. (1999). School inputs and educational outcomes in South Africa. *Quarterly Journal of Economics*, 1047–1084.

Cattaneo, M. D., Frandsen, B. R., & Titiunik, R. (2015). Randomization inference in the regression discontinuity design: An application to party advantages in the US Senate. *Journal of Causal Inference*, *3*(1), 1-24.

Center for Evaluation and Assessment (CEA). (2008). *PIRLS* (Progress in International Reading Literacy Study) *2006 Summary Report*. Pretoria: University of Pretoria.

Centre for Development and Enterprise (CDE). (2010a). *The maths and science performance of South Africa's public schools: some lessons from the past decade*. Johannesburg: Centre for Development Enterprise.

Centre for Development and Enterprise (CDE). (2010b). *Options for talented learners from disadvantaged backgrounds*. Johannesburg: Centre for Development Enterprise.

Centre for Development and Enterprise (CDE). (2011). Schooling Reform is Possible: Lessons for South Africa from International Experience. (Round Table No. 18). Johannesburg: Centre for Development Enterprise.

Centre for Education Policy Development & Joint Education Trust (2009). *Final evaluation of the foundation phase Teacher Development Project of the Zenex Foundation* (Report). Johannesburg: CEPD & JET.

Chambers, J., & Parrish, T. (1994a). Developing a resource cost database. In W. S. Barnett (Ed.), *Cost analysis for education decisions: Methods and examples* (Vol. 4, pp.23-44). Greenwich, CT: JAI.

Chambers, J., and Parrish, T. (1994b). Modeling resource costs. In W. S. Barnett (Ed.), *Cost analysis for education decisions: Methods and examples* (Vol. 4, pp. 7-21). Greenwich, CT: JAI.

Chamber of Mines of South Africa, *Facts and Figures 2012*, Johannesburg: Chamber of Mines of South Africa, 2013, http://www.bullion.org.za/documents/Chamber%20facts%20and%20figures%202013.pdf.

Chandler, J., Clarke, M., McKenzie, J., Boutron, I.,& Welch, V. (Eds.) (2013). Cochrane methods. *Cochrane Database of Systematic Reviews* 2013 (Suppl 1), pp. 1-58. Chicester :Wiley.

Chen, H.T. (1990) *Theory-driven evaluations.* Thousand Oaks, CA: SAGE Publications.

Chen, H.-T.(1994).Theory-driven evaluations: Need, difficulties, and options. *Evaluation Practice*, *15* (1), 79–82.

Chen, H.T. (2005). *Practical program evaluation: assess and improve program planning, implementation, and effectiveness*. Thousand Oaks, CA: Sage.

Chen, H. T., & Rossi, P. H. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review*, *7*, 283–302`.

Chen, H., & Rossi, P.H. (1987). The theory-driven approach to validity. *Evaluation and Program Planning, 10*, 95-103.

Chenery, H. B., & Strout, A. M. (1966). Foreign assistance and economic development. *The American Economic Review*, *56*(4), 679-733.

Chisholm, L. (2003). *The politics of curriculum review and revision in South Africa*. Paper presented at the Oxford International Conference on Education and Development, 9-11 September 2003.

Chisholm, L. (Ed.) (2004). *Changing class: Education and social change in post-apartheid South Africa*. Cape Town: HSRC Press.

Chisholm, L., Hoadley, U., Kivilu, M. (2005). *Educator workload in South Africa. Report prepared for the Educator Labour Relations Council.* Pretoria: HSRC

Chisholm, L., Volmink, J., Ndhlovu, T., Potenza, E., Mahomed, H., Muller, J., & Malan, B. (2005). *A South African curriculum for the twenty-first century* (Report of the Review Committee on Curriculum), Pretoria: HSRC

Chisholm, L., & Bagele, C. (2012). Contexts of educational policy change in Botswana and South Africa. *Prospects* 42.4 (2012): 371-388.

Christie, P. (1999). OBE and the unfolding policy trajectories: Lessons to be learned, In J. Jansen and P. Christie (eds.), *Changing curriculum: Studies on outcomes-based education in South Africa.* Kenwyn: Juta.

Christie, C. A., Ross, R. M., & Klein, B. M. (2004). Moving toward collaboration by creating a participatory internal-external evaluation team: A case study. *Studies in Educational Evaluation*, *30*(2), 125-134.

Christie, P., Butler, D., & Potterton, M. (2007). *Schools that work* (Report to the Minister of Education). Pretoria: Department of Education, Republic of South Africa.

CIDA (2008). *Presentation of the key results-based management terms and definitions for the 2008 results-based management policy statement*. Performance Management Division, Strategic Policy and Performance Branch. Ottawa: CIDA

CIDA (2009). *RBM tools at CIDA: How to guide*. *Instruction for the use of CIDA's three main results-based management (RBM) working tools; the logic model, performance measurement framework and risk register*. Ottawa,: CIDA.

CIDA (2012). CIDA learns: lessons from evaluations (Annual Report 2011-2012). Quebec: CIDA

Clark, N.L., & Worger, W. H. (2003). *South Africa: The rise and fall of apartheid.* Harlow: Longman.

Clemens, M. A., Radelet, S., & Bhavnani, R. R. (2004). *Counting chickens when they hatch: the short term effect of aid on growth* (Center for Global Development Working Paper No. 44). Washington,D.C.: Center for Global Development.

Cochrane collaboration (2014). Cochrane qualitative and implementation methods group. Retrieved on 27/10/2014 from http://cqim.cochrane.org.

Coe, R. (2002). *It's the effect size, stupid: what effect size is and why it is important*, available at http://www.leeds.ac.uk/educol/documents/00002182.htm

Coelli, M., & Green, D. A. (2012). Leadership effects: School principals and student outcomes. *Economics of Education Review*, *31*(1), 92-109.

Coetzee, M., & van der Berg, S. (2012). *Report on a scoping study to evaluate the feasability of undertaking an impact evaluation of the national school feeding programme and Grade R*, based on existing data sources (Study for the Department of Basic Education and the Department of Policy Evaluation and Monitoring in the Presidency). Stellenbosch: Stellenbosch University.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale: Lawrence Eribaum.

Collier, P., & Dollar, D. (2002). Aid allocation and poverty reduction. *European Economic Review*, *46*(8), 1475-1500.

Collier, P. (2008). *The bottom billion: Why the poorest countries are failing and what can be done about it*. New York, NY: Oxford University Press.

Conley-Tyler, M. (2005). A fundamental choice: internal or external evaluation? *Evaluation Journal of Australasia*, *4*(1/2), 3.

Conn, K. M. (2014). *Identifying effective education interventions in Sub-Saharan Africa: a meta-analysis of rigorous impact evaluations*. Unpublished PhD thesis, Columbia University, New York

Cook, T. D., Campbell, D. T., & Day, A. (1979). *Quasi-experimentation: design & analysis issues for field settings* (Vol. 351). Boston: Houghton Mifflin Company.

Cook, T. D., & Campbell, D. T. (1986). The causal assumptions of quasi-experimental practice. *Synthese*, *68*(1), 141-180.

Cooke, B., & Kothari, U. (2001). The case for participation as tyranny. In B. Cooke, & U. Kothari (Eds.), *Participation: The New Tyranny* (pp. 1–15). London: Zed Books.

Cornell, J. E. & Mulrow, C. D. (1999). Meta-analysis. In: H. J. Adèr & G. J. Mellenbergh (Eds). *Research Methodology in the social, behavioral and life sciences* (pp. 285–323). London: Sage.

Corruption Watch (2012). *Unpacking SA's education crisis: Part six*. Retrieved from http://www.corruptionwatch.org.za/unpacking-sas-education-crisis-part-six/

Coryn, C. L. S. (2005). Practical program evaluation: Assessing and improving planning, implementation, and effectiveness (Review by H. T. Chen). *American Journal of Evaluation*, *26*,405-407.

Cranney A, Guyatt G, Griffith L, Wells G, Tugwell P, Rosen C, et al. (2002). Summary of meta-analyses of therapies for post-menopausal osteoporosis. *Endocrine Review*, 23, 570

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into practice*, *39*(3), 124-130.

Croft, A. (2013). Promoting access to education for disabled children in low-income countries. Do we need to know how many disabled children there are? *International Journal of Educational Development, 33*(3): 233-243.

Crouch, L., & Mabogoane, T. (2001). No magic bullets, just tracer bullets: The role of learning resources, social advantage, and education management in improving the performance of South African schools. *Social Dynamics*, *27*(1), 60-78.

Cummings, T. G., & Worley, C. G. (2014). *Organization development and change*. Cengage learning. Stamford: Cengage Learning

Cummins, J. (2009). Multilingualism in the english-language classroom: Pedagogical considerations. *TESOL Quarterly 43*(2), 317-321.

Davies, P. (2010). *A review of the roles and activities of new development partners*. (CFP Working Paper Series No. 4). Washington, DC: World Bank

Davies, P. (2012). The state of evidence-based policy evaluation and its role in policy formation. *National Institute Economic Review*, 219(1), R41-R52.

Davies, P. (2012). *Introduction to impact evaluation*. CLEAR/3ie workshop for policy-makers held at University of Witwatersrand, Johannesburg, September 2012.

Davies, P (2015). *The challenges of evidence synthesis*, Presentation at CLEAR Eval-Africa Seminar, Johannesburg, 3 June 2015.

De Chaisemartin, T. (2010). *Evaluation of the cape teaching and leadership institute*. Johannesburg: JET Education Services. Available at www. jet. org. za.

De Clercq, F. (2008). Teacher quality, appraisal and development: the flaws of the IQMS. *Perspectives in Education*, *26*(1),7–18.

De Groof, J., & Bray, W. (1996). *Education under the new Constitution in South Africa*. Leeuven: ACCO.

Deacon, & Simkins, C. (2011). *Value in the classroom: The quantity and quality of South Africa's teachers*. Johannesburg: Centre for Development and Enterprise (CDE).

Dearden, L., Reed, H., & van Reenen, J. (2006). The impact of training on productivity and wages: Evidence from British Panel Data. *Oxford Bulletin of Economics and Statistics*, *68* (4), 397-421.

Dearden, L., Emmerson, C., Frayne, C., & Meghir, C. (2008). *Education subsidies and school drop-out rates*. London: Institute for Fiscal studies.

Deaton, A. (2009). *Instruments of development: randomization in the tropics, and the search for the elusive keys to economic development* (NBER Working Paper No. 14690). Cambridge, MA: National Bureau of Economic Research (NBER).

Deeks, J. J., Dinnes, J., D'amico, R., Sowden, A. J., Sakarovitch, C., Song, F., Petticrew, M., & Altman, D. G. (2003). Evaluating non-randomised intervention studies. *Health Technology Assessment* , *7*(27), iii-x.

Dehejia, R.H., & Wahba,S. (2002). Propensity score matching methods for nonexperimental causal studies. *The Review of Economics and Statistics*, *84*(1),151-161.

Dehejia, R. (2013). *The porous dialectic: experimental and non-experimental methods in development economics* (WIDER Working Paper No. 2013/011). Helsinki: WIDER (World Institute of Development Economics Research).

Dehn, J., Reinikka, R., & Svensson, J. (2003). Survey tools for assessing performance in service delivery. In F. Bourguignon, & L.A. Pereira da Silva (Eds.), *Evaluating the poverty and distributional impact of economic policies (techniques and tools)* (pp.170-186). Washington, D.C.: World Bank.

Denzin, N. (2006). *Sociological methods: A sourcebook* (5th ed.). Chicago, IL: Aldine Transaction.

Department for International Development (DFID) (2011). *Bilateral aid review* (Technical Report). London: DFID.

Department of Basic Education (2014). *National Norms and Standards for School Funding*. Government Gazette, Vol. 583, No. 37230, 17 January 2014. Pretoria: Government of South Africa.

Department of Basic Education (DBE). (2009). *School funding and management in South Africa.* Pretoria: DBE.

Department of Basic Education (DBE). (2010). *Annual Report 2009/2010*. Pretoria: DBE.

Department of Basic Education (DBE). (2011). *National Senior Certificate Examination*: *National Report on Learner Performance in Selected Subjects*. Pretoria: DBE.

Department of Basic Education (DBE). (2012a). *Education Statistics in South Africa 2010*. Pretoria: DBE.

Department of Basic Education (DBE). (2012b). *General household survey (GHS) 2010: Focus on schooling*. Pretoria: DBE.

Department of Basic Education (DBE). (2012c). *Report on the Annual National Assessments of 2011*. Pretoria: DBE.

Department of Basic Education (DBE). (2013). *Report on the Annual National Assessment of 2013*. Pretoria: DBE.

Department of Basic Education (DBE). (2015). *National Senior Certificate Examination*: *National Report on Learner Performance in Selected Subjects*. Pretoria: DBE.

Department of Basic Education (DBE). (2011). Curriculum and Assessment Policy Statement (CAPS). Pretoria: DBE

Department of Education (DoE). (2002). *Education Statistics in South Africa at a Glance* (Annual Report). Pretoria: DoE.

Department of Education (DoE) (2004) *Education Statistics in South Africa at a Glance 2003.* Pretoria: DoE

Department of Education (DoE) (2005) *Education Statistics in South Africa at a Glance 2004.* Pretoria: DoE

Department of Education (DoE) (2006) *Senior Certificate: Report on the 2005 examination results.* Pretoria: DoE.

Department of Education (DoE). (2007). *Education Statistics in South Africa at a Glance* (Annual Report). Pretoria: DoE

Department of Education (DoE) (2008). *Education statistics in South Africa* at a Glance Pretoria: DoE.

Department of Planning, Monitoring and Evaluation (DPME). (2010). *Guide to the Outcomes Approach* (Guide). Pretoria, South Africa: The Presidency, South Africa.

Department of Planning, Monitoring and Evaluation (DPME). (2011). *National evaluation policy framework* . Pretoria: South Africa.

Desai, R., & Kharas, H. (2008). The California Consensus: can private aid end global poverty? *Survival: Global Politics and Strategy*, *50*(4), 155-168.

Desai, R.M., & Kharas, H. (2010). Democratizing foreign aid: online philanthropy and international development assistance. *International Law and Politics*, *42*(4):1111-1142.

Development Bank of Southern Africa (DBSA). (2008). *DBSA infrastructure barometer, 2008: economic and social infrastructure in South Africa: Scenarios for the future*. Johannesburg, South Africa: DBSA.

Dhaliwal, D. S., Radhakrishnan, S., Tsang, A., & Yang, Y. G. (2012). Nonfinancial disclosure and analyst forecast accuracy: international evidence on corporate social responsibility disclosure. *The Accounting Review*, 87(3), 723-759.

Dichter, T. W. (2003). *Despite good intentions: why development assistance to the Third World has failed*. Cambridge: University of Massachusetts Press.

Dieltiens, V., & Mandipaza, B. (2015). *The carrot and the stick: School improvement through support and accountability. Twenty Years of Education Transformation in Gauteng 1994 to 2014*. Johannesburg: African Minds / Gauteng Department of Education.

Dietrich, S., & Wright, J. (2012). Foreign aid and democratic development in Africa. *UNU-WIDER Working Paper*, (2012/20).

DiNardo, J. (2008). Natural experiments and quasi-natural experiments. In S. Durlauf, & L. Blume (Eds.), *The new Palgrave dictionary of economics* (2nd ed.). Basingstoke, UK: Palgrave Macmillan.

Donaldson, S.I. (2005). Using program theory-driven evaluation science to crack the Da Vinci Code. *New Directions for Evaluation* 106, 65-84.

Donegan, S., Williamson,P., Gamble,C., & Tudur-Smith, C. (2010). Indirect comparisons: A review of reporting and methodological quality. *PLoSone, 5* (11), e11054.

Donaldson, S. (2007). *Program Theory – Driven Evaluation Science. Strategies and Applications,* New York: Lawrence Erlbaum Associates, Taylor &Francis Group.

Doucouliagos, H., & Paldam, M. (2009). The aid effectiveness literature: The sad results of 40 years of research. *Journal of Economic Surveys*, *23*(3), 433-461.

Drucker, P. F. (1954). Management by objectives and self-control. *Practice of Management*. New York: Harper & Row Publishers

Duflo, E. (2003). Grandmothers and granddaughters: old-age pensions and intra-household allocation in South Africa. *The World Bank Economic Review*, *17*(1), 1-25.

Duflo, E., Glennerster,R., & Kremer, M. (2008). Using randomization in development economics research: A toolkit. In T. Schultz & J. Strauss (eds.), *Handbook of Development Economics*, Vol. 4. Amsterdam and New York: North Holland.

Duflo, E., Dupas, P., & Kremer, M. (2008). *Peer effects and the impact of tracking: Evidence from a randomized evaluation in Kenya*. Cambridge: National Bureau of Economic Research

Duflo, E., Glennerster,R., & Banerjee,A (2009). Abdul Latif Jameel Poverty Action Lab executive training: *Evaluating social programs*, Spring 2009. Retrieved from http://ocw.mit.edu/resources/res-14-001-abdul-latif-jameel-poverty-action-lab-executive-training-evaluating-social-programs-2009-spring-2009/

Duflo, E., Dupas, P., Kremer, M., & Kenya, C. (2012). *School governance, teacher incentives, and pupil-teacher ratios: experimental evidence from Kenyan primary schools.* NBER Working Paper N. 17939. Cambridge: National Bureau of Economic Research

Dunning, T.(2008). Improving causal inference: strengths and limitations of natural experiments. *Political Science Quarterly, 61* (2), 282–293.

Du Toit, J. L., & Mouton, J. (2013). A typology of designs for social research in the built environment. *International Journal of Social Research Methodology*, *16*(2), 125-139.

Duvendack, M., Hombrados, J. G., Palmer-Jones, R., & Waddington, H. (2012). Assessing 'what works' in international development: meta-analysis for sophisticated dummies. *Journal of Development Effectiveness*, 4(3), 456-471.

Easterly, W. (2006). Reliving the 1950s: the big push, poverty traps, and takeoffs in economic development. *Journal of Economic Growth*, *11*(4), 289-318.

Easterly, W. (2007). *The white man's burden: why the West's efforts to aid the rest have done so much ill and so little good*. New York: Penguin Press

Easterly, W. (2007). Was development assistance a mistake? *The American Economic Review* , *97* (2), 328-332 .

Easterly, W., & Levine, R. (2003). Tropics, germs, and crops: How endowments influence economic development (NBER Working Paper 9106).Cambridge,: NBER (National Bureau of Economic Research).

Eisenbruch, M. B. (2015). I., & Harris, M.(2005). *Results based accountability*. Sydney: Centre for Culture and Health, University of New South Wales

Elmore, R. (2008). Leadership as the practice of improvement. *Improving School Leadership*, 2, 40-41.

Ellis, P. (2010). *The Essential Guide to Effect Sizes: An Introduction to Statistical Power, Meta-Analysis and the Interpretation of Research Results*. United Kingdom: Cambridge University Press.

Ellis, T. J., & Levy, Y. (2009). Towards a guide for novice researchers on research methodology: Review and proposed methods. *Issues in Informing Science and Information Technology*, *6*, 323-337.

Epstein, J. L. (2001). *School, family, and community partnerships: preparing educators and improving schools*. Boulder: Westview Press

Evans, D., & Popova, A. (2014). *Cost-effectiveness measurement in development: accounting for local costs and noisy impacts.* World Bank Policy Research Working Paper, (7027). Washington DC: World Bank

Evans, D., Kremmer, M., & Ngatia, M. (2014). *Schooling costs, school participation, and long-run outcomes: evidence from Kenya.* Unpublished Working Paper

Evans, D., & Popova, A. (2015). *What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews. An Analysis of Divergent Findings in Systematic Reviews* (February 26, 2015). World Bank Policy Research Working Paper, (7203). Washington DC: World Bank

Evans, S., & Erasmus, J. (2012). *Textbook crisis: Education department favoured dodgy tender*. Mail and Guardian (29 June 2012). Retrieved from http://mg.co.za/article/2012-06-29-education-department-favoured-dodgy-tender

Farrington, D. P., & Welsh, B. C. (2005). Randomized experiments in criminology: What have we learned in the last two decades?. *Journal of Experimental Criminology*, *1*(1), 9-38.

Field, A. (2009). *Discovering statistics using SPSS* (3rd edition). London, UK: Sage Publications.

Fiske, E.B., & Ladd, H. F. (2004). Balancing public and private resources for basic education: School fees in post-apartheid South Africa. In L. Chisholm (Ed.), *Changing class: Education and social change in post-apartheid South Africa*. Cape Town: HSRC Press.

Fleisch B. (2002). *A case study of the Gauteng Department of Education's Education Action Zones,* Research on School Development Study. Johannesburg: JET Education Services.

Fleisch, B., & Woolman, S. (2004). On the constitutionality of school fees: A reply to Roithmayr. *Perspectives in Education*, *22*(1), 111-123.

Fleisch, B. (2006). Bureaucratic accountability in the education action zones of South Africa. *South African Journal of Education*, *26*(3), 369–382.

Fleisch, B. (2008). *Primary education in crisis: Why South African school children underachieve in reading and mathematics*. Johannesburg: Juta & Co. Ltd.

Fleisch, B., Taylor, N., Herholdt, R., & Sapire, I. (2011). Evaluation of the Back to Basics Mathematics Workbooks: A randomized Controlled Trial of the primary mathematic research project. *South African Journal of Education*, *31*(4), 488–501.

Fleisch, B., & Schöer, V. (2014). Large-scale instructional reform in the Global South: insights from the mid-point evaluation of the Gauteng Primary Language and Mathematics Strategy. *South African Journal of Education*, *34*(3), 01–12.

Fleisch, B., Taylor, S., Schöer, V., & Mabogoane, T. (2015). *Assessing the impact of the RCUP: The findings of the impact evaluation of the Reading Catch-Up Programme*. Report, Johannesburg

Flint, M., Cameron, C., Henderson, S., Jones, S., & Ticehurst, D. (2002). *Development Effectiveness Report 2001.* London: DFID.

Fosu, A. K. (2009). Inequality and the impact of growth on poverty: comparative evidence for sub-Saharan Africa. *Journal of Development Studies*, *45*(5), 726-745.

Fowler, A. (2013). *Striking a balance: A guide to enhancing the effectiveness of non-governmental organisations in international development*. Routledge.

Friedman, M. (2002). *Capitalism and freedom*. Chicago: University of Chicago Press.

Friedman, M. (2005). *Trying hard is not good enough*. Bloomington: Trafford Publishers

Friedman, W., Kremer, M., Miguel, E., & Thornton, R. (2011). *Education as liberation*? (NBER Working Paper No. 16939). Cambridge, MA: NBER (National Bureau of Economic Research).

Funnell, S.C., & Rogers, P. J. (2011). *Purposeful program theory: Effective use of theories of change and logic models*. San Francisco, CA: Jossey-Bass/Wiley.

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L.B., & Vermeersch, C.M. J.(2011). *Impact evaluation in practice.* Washington, D.C.: World Bank.

Giamporcaro, S., & Dhlamini, X. (2015). *The african investing for impact barometer* (3rd edn.). Cape Town: Bertha Center for Social Innovation and Entrepreneurship, Graduate School of Business, UCT.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*(10), 3-8.

Glass, G.V., McGraw, B. & Smith, M.L. (1981). *Meta-analysis in social research*. Beverly Hills: Sage.

Glennerster, H. (2009). *Understanding the finance of welfare*. Bristol: Policy Press.

Glennie, J. (2011). *The role of aid to middle-income countries* (ODI Working Paper). London: ODI. Retrieved from http://www.odi.org.uk/sites/odi.org.uk/files/odi-assets/publications- opinion-files/7189.pdf

Glewwe, P., Kremer, M., Moulin, S., & Zitzewitz, E. (2004). Retrospective vs. prospective analyses of school inputs: The case of flip charts in Kenya. *Journal of Development Economics*, 74(1), 251-268.

Glewwe, P., & Kremer, M. (2006). Schools, teachers, and education outcomes in developing countries. *Handbook of the Economics of Education*, 2, 945–1017.

Glewwe, P; & Miguel, E.A. (2007). The impact of child health and nutrition on education in less developed countries. *In Handbook of development economics*. *4*, 3561-3606.

Glewwe, P. W., Hanushek, E. A., Humpage, S. D., & Ravina, R. (2011). *School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010* (No. w17554). National Bureau of Economic Research.

Glewwe, P. (2012). How much of observed economic mobility is measurement error? IV methods to reduce measurement error bias, with an application to Vietnam. *The World Bank Economic Review*, *26*(2), 236-264.

Gough, D., Oliver, S., & Thomas, J. (Eds.). (2012). *An introduction to systematic reviews*. London: Sage.

Goldberger, A. S. (1964). *Econometric theory.* New York: John Wiley & Sons.

Gosling, L., & Edwards, M. (2003). *Toolkits: A practical guide to planning, monitoring, evaluation and impact assessment*. London: Save the Children UK.

Govender, L. (2004). Teacher unions, policy struggles and educational change, 1994 to 2004. *Changing class: education and social change in post-apartheid South Africa*, 267-291.

Grant-Lewis, S. & Motala, S. (2004). Education decentralization and the quest for equality, quality and democracy. In L. Chisholm (ed.) *Changing class: education and social change in post-apartheid South Africa*. Pretoria: HSRC Press.

Greenhill, R., Prizzon, A., & Rogerson, A. (2013). *The age of choice: developing countries in the new aid landscape* (Synthesis Report). London: ODI.

Griffin, K. B., & Enos, J. L. (1970). Foreign assistance: objectives and consequences. *Economic development and cultural change*, *18*(3), 313-327.

Grissom, J. A., Kalogrides, D., & Loeb, S. (2014). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, 0162373714523831.

Guillaumont, P. & Chauvet, L. (2001). Aid and performance: A reassessment. *Journal of Development Studies*, *37*(6), 66-92.

Gustafsson, M. (2007). Using the hierarchical linear model to understand school production in South Africa. *South African Journal of Economics*, *75*(1), 84-98.

Gustafsson, M., Taylor, S., & others. (2013). *Treating schools to a new administration. The impact of South Africa's 2005 provincial boundary changes on school performance*. Stellenbosch: Stellenbosch University.

Gwilliam, K. (1997). Sustainable transport and economic development. *Journal of Transport Economics and Policy*, *31*(3), 325-330.

Habib, A. (2013). Opening remarks made at guest lecture by Kgalema Motlanthe, Deputy President of South Africa at the Wits School of Governance, Johannesburg in September 2013.

Hamann, R. (2004). Corporate social responsibility, partnerships, and institutional change: The case of mining companies in South Africa. *Natural Resources Forum 28*, 278-290.

Hansen, L. O., Young, R. S., Hinami, K., Leung, A., & Williams, M. V. (2011). Interventions to reduce 30-day rehospitalization: a systematic review. *Annals of Internal Medicine*, 155(8), 520-528.

Hanushek, E., & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, *17*(4), 267–321.

Haq, M. (1995). The advent of the human development report. In ul Haq, M. (Ed.), *Reflections on human development* (pp. 24-45). Oxford: Oxford University Press.

Harley, K. ,& Wedekind, V. (2004). Political change, curriculum change and social formation, 1990 to 2002. In L. Chisholm (Ed.). *Changing class: Education and social change in post-apartheid South Africa* (pp. 195-220). Cape Town: Compress.

Hartmann, D. P., Gottman, J. M., Jones, R. R., Gardner, W., Kazdin, A. E., & Vaught, R. S. (1980). Interrupted time-series analysis and its application to behavioral data. *Journal of Applied Behavior Analysis*, *13*(4), 543-559.

Harris, D. (2009). Toward policy-relevant benchmarks for interpreting effect sizes combining effects with costs. *Educational Evaluation and Policy Analysis*, 31(1), 3-29.

Hattie, J. (2009). *Visible learning: A synthesis of 800+ meta-analyses on achievement*. Abingdon: Routledge.

Hattingh, A. (2003). Examining learning achievement and experiences of science learners in a problem-based learning environment. *South African Journal of Education*, 23(1), p–52.

Haveman, R., & Wolfe, B. (1995). *The "window problem" in studies of children's attainments: A methodological exploration* (Institute for Research on Poverty Discussion Papers). Madison: University of Wisconsin Institute for Research on Poverty.

Heckman, J. J. (1979). *Statistical models for discrete panel data*. Chicago: Department of Economics and Graduate School of Business, University of Chicago.

Heckman, J. J. (1995). *Randomization as an instrumental variable* (NBER Technical Working Paper No. 184). Cambridge: NBER.

Heckman, J. J., Ichimura, H., & Todd, P. (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies*, 65(2), 261-294.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics*, 6(2), 107-128.

Hedges, L. V. (1982). Fitting continuous models to effect size data. *Journal of Educational and Behavioral Statistics*, 7(4), 245-270.

Hedges, L. V., & Olkin, I. (2014). *Statistical methods for meta-analysis*. Orlando: Academic press.

Hicks, D.A. (1997). The inequality-adjusted human development index: a constructive proposal. *World Development*, *25*(8), 1283-1298.

Hickey, S., & Mohan, G. (2004). *Participation: from tyranny to transformation? Exploring new approaches to participation in development*. London: Zed books.

Higgins, J. P. T., & Green, S. (Eds.) (2012). *Cochrane handbook for systematic reviews of interventions* (Version 5.1. 0). Chichester: Wiley-Blackwell.

Hellman, L. (2012). *GPLMS Intersen Catchup Programme: analysis of results* (Main Report). Johannesburg: Class Act Educational Services.

Hobden, P. & Hobden, S. (2009). RUMEP Bed Project (Summative Evaluation). Johannesburg: Zenex Foundation.

Hobden, P., & Hobden, S. (2009). *Quality Project in Education* (QPiE). Johannesburg: QPiE and Schaffer & Associates.

Hollander, S. (1998). The canonical classical growth model: content, adherence and priority. *Journal of the History of Economic Thought*, *20*, 253-77.

Hombrados, J., & Waddington, H, (2012), *Including quasi-experiments in meta-analysis*, presentation at 3ie/Campbell Collaboration Systematic Review Symposium in Dhaka (Bangladesh), December 2012.

Hombrados, J. G., & Waddington, H. (2012). *Internal validity in social experiments and quasi-experiments: An assessment tool for reviewers*. Mimeo, London: 3ie (International Initiative for Impact Evaluation).

Howie, S.J., Venter, E., van Staden, S., Zimmerman, L., Long, C., Scherman, V., & Archer, E. (2007). *PIRLS 2006 summary report: South African children's reading literacy achievement*. Pretoria: Centre or Evaluation and Assessment.

Howes, S., Otor, S., & Rogers, C. (2011). *Is there a micro-macro paradox in international aid, or do the data deceive?* Canberra: Center for Development Policy, Australian National University.

Hudson Institute Center for Global Prosperity (2013). *The index of global philanthropy and remittances with a special report on emerging economies.* Washington, D.C.: Hudson Institute.

Hunt, K., Schoer, V., Ntuli, M., Rankin, N., & Sebastiao, C. (2011) A blurred signal? Comparability of NSC mathematics scores and former SC mathematics scores: how consistent is the signal across time? *Education as Change*, *15* (1), 3-16.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: cumulating research findings across studies* (Vol. 4). Sage Publications, Inc.

Iglehart, J. K. (2009). Prioritizing comparative-effectiveness research: IOM recommendations (Perspective). *New England Journal of Medicine*, *361*(4), 325-328.

Jack, W., & Lewis, M. (2009). *Health investments and economic growth: macroeconomic evidence and microeconomic foundations*. Washington DC: World Bank Policy Research Working Paper Series, Vol.(2009)

Jacoby, H. G. (2002). Is there an intrahousehold 'flypaper effect'? Evidence from a school feeding programme. *The Economic Journal*, 112(476), 196-221.

Jacobs, I. (1999). Strategic priorities for the GDE for the next five years. *Press Release: Gauteng Department of Education*.

Jalan, J., & Ravallion, M. (2003). Estimating the benefit incidence of an antipoverty program by propensity score matching. *Journal of Business & Economic Statistics*, 21(1), 19-30.

Jansen, J (2004a). *Accounting for autonomy*. The 41st TB Davie Memorial Lecture [delivered at the] University of Cape Town, 26 August 2004. Available at *http://www.chet.org.za/issues/41stTBDavie.doc* (accessed 10 May 2006).

Jansen, J (2004b). *When does a university cease to exist?* The 40th Hoernlé Memorial Lecture, South African Institute of Race Relations, 17 November 2004. Available at http://www.sanpad.org.za/Research/Jansen%20-%20Hoernle%20lecture.doc.

Jansen, J. (1999). *The school curriculum since apartheid: intersections of politics and policy in the South African transition.* Pretoria: University of Pretoria.

Jansen, J.D. (1997). Essential alterations? A critical analysis of the states syllabus revision process. *Perspectives in Education*, *17*(2),1-11.

Jepma, C. J. (1991). *The tying of aid*. Paris: OECD Publishing.

JET Education Services. (2007). *Towards optimal usage of mindset resources in South African Schools: Five Case Studies in Western Cape and Gauteng*. Johannesburg: JET..

JET Education Services. (2009). *Evaluation of the CPD programme for curriculum advisers and educators in the Limpopo Province over 2008.* Johannesburg: Joint Education Trust.

Jick, T. D. (1979). Mixing qualitative and quantitative methods: triangulation in action. *Administrative Science Quarterly, 24*, 602-611.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: a research paradigm whose time has come. *Educational Researcher*, *33*(7), 14-26.

Johnson, J., & Lamdany, R. (2005). *OED methodology for country assistance evaluations.* Washington, D.C.: OED, World Bank.

Johnstone, R. (1976). *Class, race and gold*. London: Routledge.

Joint Education Trust (JET). (2007). *Reflections on Khanyisa 2003-2007*. Polokwane: Limpopo Department of Education

Joint Education Trust (JET). (2009). *Khanyisa School Transformation Programme*. Johannesburg: JET Education Services

Joint Education Trust (JET). (2016). *History-JET Education Services*. Retrieved on 27 January 2016 from http://www.jet.org.za/about-jet/history

Jüni, P., Altman, D. G., & Egger, M. (2001). Assessing the quality of a randomised controlled trial. In M. Egger, G. Davey Smith, & D.G. Altman (Eds.), *Systematic reviews in health care: meta-analysis in context* (2nd ed., pp.87-108). London: BMJ Publishing Group.

Kanjee A. (2007). *Large-scale assessments in South Africa: supporting formative assessments in schools*. Paper presented at the 33rd International Association for Educational Assessment Conference, Baku, Azerbaijan, 16 to 21 September.

Kanjee, A. (2012). Academic achievement in South Africa. In J. Hattie & E. M. Anderman (Eds.). *International guide to student achievement* (Chapter 9.6). New York and London: Routledge.

Karlan, D., & Zinman, J. (2010). Expanding credit access: using randomized supply decisions to estimate the impacts. *Review of Financial Studies*, *23,*1, 433–464.

Kaphesi, E. (2001). *Effects of home language on pupils' performance in mathematics*. London: British Society for Research into Learning Mathematics

Kellaghan, T., & Greaney, V. (2001). *Using assessment to improve the quality of education.* Paris: International Institute for Educational Planning.

Kharas, H. & Linn, J.F. (2008). *Better aid: responding to gaps in effectiveness* (Policy Brief 2008-06, Brookings global economy and development: Global views). Washington, D.C.: Brookings Institution.

Khandker, S., Koolwal, B., & Samad, H. (2009). *Handbook on impact evaluation: Quantitative methods and practices*. Washington,D.C.: World Bank.

Kidder, L. H., & Fine, M. (1987). Qualitative and quantitative methods: when stories converge. *New directions for program evaluation*, *1987*(35), 57-75.

Kindornay, S., & Morton, B. (2009). *Development effectiveness: towards new understandings* (Issues Brief, Development Cooperation Series). Ottawa, Canada: The North-South Institute.

Kindornay, S., & Morton, B. (2009a). *From aid to development effectiveness: A Working Paper.* Ottawa: The North-South Institute.

Kmenta, J., & Ramsey, J. B. (Eds.). (2014). *Evaluation of econometric models*. New York: Academic Press.

Knack, S. (2001). Aid dependence and the quality of governance: cross-country empirical tests. *Southern Economic Journal*, 310-329.

Kotvojs, F. (2006). *Contribution analysis: A new approach to evaluation in international development*. Paper presented at the International Conference of the Australasian Evaluation Society, Darwin, Australia. Retrieved from http://www.aes.asn.au/conferences/2006/papers/ 022%20Fiona%20Kotvojs.pdf

Kremer, M., Brannen, C., & Glennerster, R. (2013). The challenge of education and learning in the developing world. *Science*, *340*(6130), 297–300.

Kremer, M., & Glennerster, R. (2011). Improving Health in Developing Countries. *Handbook of Health Economics*, *2*, 201-315.

Krishnaratne, S., White, H., & Carpenter, E. (2013). *Quality education for all children? What works in education in developing countries.* Working Paper (20). New Delhi: International Initiative for Impact Evaluation (3ie)

Kruse, S.E. (1998). *Hvilke resultater skaper de frivillige organisasjonene? Erfaringer fra et utvalg evalueringer*. Oslo: NORAD (Norwegian Agency for Development Cooperation).

Kruse, S.-E. (2003). *Searching for impact and methods: Meta-evaluations of NGO experience*. Paper presented at the Fifth Biennial World Bank Conference on Evaluation and Development (Evaluating development effectiveness: Challenges and the way forward), Washington, D.C., 15-16 July 2003.

Kusek, J. Z., & Rist, R. C. (2004). *Ten steps to a results-based monitoring and evaluation system: A handbook for development practitioners*. Washington, D.C.: World Bank Publications.

Kusek, J.Z., & Khatouri, M. (2006). *Results-based monitoring and evaluation in bank projects*. Washington, D.C.: World Bank.

Lahey, R., & Nielsen, S. B. (2013). Rethinking the Relationship Among Monitoring, Evaluation, and Results-Based Management: Observations From Canada. *New Directions for Evaluation*, *2013*(137), 45-56.

Landsberg, E., Kruger, D., & Nel, N. (Eds.) (2005). *Addressing barriers to learning: A South African perspective.* Pretoria: Van Schaik.

Laxminarayan, R., Chow, J., & Shahid-Salles, S. A. (2006). Intervention cost-effectiveness: overview of main messages. In DT Jamison, JG Breman, AR Measham, G Alleyne, M Claeson, D B Evans, P Jha, A Mills, & P Musgrove (Eds.), *Disease control priorities in developing countries* (2nd ed., Chapter 2). Washington, D.C.: World Bank

Lee, C.J., Park, H.J., & Lee, H. (2009). Shadow education systems. In G. Sykes, B. Schneirder, & D.N. Plank (Eds). *Handbook of Education Policy Research*. (pp. 901-919) New York: Routledge for the American Educational Research Association.

Lee, D.S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48* (2), 281-355.

Leeuw, F. L., & Cooksy, L. J. (2005). Evaluating the performance of development agencies: The role of metaevaluations. *World Bank series on evaluation and development*, *7*, 95-108.

Leeuw, F., & Vaessen, J. (2009). *Impact evaluations and development* (NONIE Guidance on Impact Evaluation). Washington, DC: IEG (Independent Evaluation Group).

Leithwood, K., Seashore Louis, K., Anderson, S., & Wahlstrom, K. (2004). Executive Summary: Review of Research: How Leadership Influences Student Learning. Minneapolis: University of Minnesota

Levin, H.M. (1975). Cost-effectiveness analysis in evaluation research. In M. Guttentag, & E. Struening (Eds.), *Handbook of evaluation research* (Vol. 2, pp. 89-122). Beverly Hills, CA: Sage Publications.

Levin, H. M., & McEwan, P. J. (2001). *Cost-effectiveness analysis: methods and applications* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Levison-Johnson, J., Dewey, J., & Wandersman, A. (2009). *Getting to outcomes in systems of care:10 steps for achieving results-based accountability*. Anatnta, GA: ICF Macro.

Levy, V. (1988). Aid and growth in Sub-Saharan Africa: the recent experience. *European Economic Review*, *32*(9), 1777-1795.

Lewis, S. G., & Motala, S. (2003). Educational de/centralisation and the quest for equity, democracy and quality. In L. Chisholm. *Ten years after: Education and social change in South Africa*. Pretoria: HSRC Press.

Lindahl, C., & Catterson, J. (2005). What causes aid dependency? Evaluating thirty years of Swedish bilateral aid to Tanzania1. *Evaluating Development Effectiveness*, *7*, 133.

Lindelöw, M., & Wagstaff, A. (2003). *Health facility surveys: an introduction* (Vol. 2953). Washington DC: World Bank Publications.

Lipscomb, S., Chiang, H., & Gill, B. (2012). Value-added estimates for phase 1 of the Pennsylvania Teacher and Principal Evaluation Pilot. Full report. *Mathematica Policy Research, Inc.*

Lipsey, M. W. (1990). *Design sensitivity: statistical power for experimental research* (Vol. 19). Sage.

Lipsey, M.W. (1993). Theory as method: small theories of treatments. *New Directions for Program Evaluation*, *57*, 5-38.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: confirmation from meta-analysis. *American Psychologist*, *48*(12), 1181.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications, Inc.

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Busick, M. D. (2012). *Translating the Statistical Representation of the Effects of Education Interventions into More Readily Interpretable Forms.* (No. ED-IES-09-C-0021). Washington DC.

Louw, J., Muller, J., & Tredoux, M. (2008). Time-on-task, technology and mathematics achievement, *Evaluation and Program Planning*. 31(1), 41-50.

Lu, G., & Ades, A. E. (2006). Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association*, 101 (474), 447 - 459.

Lucas, A. M., McEwan, P. J., Ngware, M., & Oketch, M. (2014). Improving early-grade literacy in East Africa: experimental evidence from Kenya and Uganda. *Journal of Policy Analysis and Management*, *33*(4), 950-976.

Lundsgaarde, E. (2013): *Complementarity in development: bringing private foundations on board* (Briefing Paper 10/2013, German Development Institute). Retrieved from http://www.die-gdi.de/CMSHomepage/openwebcms3.nsf/(ynDK_contentByKey)/ANES- 96WDRM?Open

Luxomo, G., & Motala, S. (2012). The return home: parental involvement in children's education. In S Motala, V. Dieltiens, Y Sayed (2012) *Finding Place and Keeping Pace*: *Exploring Meaningful and Equitable Learning in South African Schools*. Pretoria: HSRC Press

Mabe, J. (2013). *Education review presentation (2009–2012): education & skills development*. Presentation at the Anglo Platinum CED review meeting, June 2013, Johannesburg.

Macdonald, K., Barrera-Osorio, F., Guaqueta,J., Patrinos, H.A., & Porta, E. (2010). *The determinants of wealth and gender inequity in cognitive skills in Latin America* (Policy Research Working Paper no. WPS5189). Washington, D.C.: World Bank.

Mahajan, V. (2011). *Africa rising: How 900 million African consumers offer more than you think*. New Jersey: Pearson Education, Inc.

Malamud, O., & Pop-Eleches,C. (2010). General education versus vocational training: Evidence from an economy in transition. *The Review of Economics and Statistics*, *92*(1),43-60.

Mandela, N. (1994). *Long walk to freedom*: the autobiography of Nelson Mandela. Randburg, South Africa: Macdonald Purnell.

Maringe, F., Masinire, A., Nkambule, T. (2015). Distinctive features of schools in multiple deprived communities in South Africa: implications for policy and leadership, *Educational Management Administration & Leadership*, Vol. 43 (3), 363-385

Marris, R. (1970). Can we measure the need for development assistance? *The Economic Journal*, 650-668.

Marneweck L (2007). *Do we need inspectorate?* Seminar Series on Making a Difference in Public Schooling. Johannesburg: Education Policy Development

Masud, N., & Yontcheva, B. (2005). *Does foreign aid reduce poverty? Empirical evidence from nongovernmental and bilateral aid* (IMF Working Paper, WP/05/100). Washington, D.C.: IMF.

Mayne, J. (2008). *Contribution analysis: An approach to exploring cause and effect* (ILAC Methodological Brief). Retrieved from http://www.cgiar-ilac.org/files/ILAC_Brief16_Contribution_Analysis_0.pdf

Maren, M. (2002). *The road to hell: The ravaging effects of foreign aid and international charity*. New York: Free Press.

Mbokazi, Z. (2015). Dimensions of successful leadership in Soweto township secondary schools, *Educational Management Administration & Leadership*, Vol. 43 (3), 468-482

McEwan, P. J. (2015). Improving learning in primary schools of developing countries: a meta-analysis of randomized experiments. *Review of Educational Research*, 85(3), 353-394.

Meinzen-Dick, R.S. (2004). How do we evaluate sustainability? In G.K. Pitman, O.N. Feinstein, & G. K. Ingram (Eds.), *Evaluating development effectiveness* (World Bank Series on Evaluation and Development, Series No.7). New Brunswick, NJ :Transaction Publishers.

Meredith, M., & Perkoski, E. (2015). Regression Discontinuity Design. *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*. 1–16.

Mertens, D. M. (2014). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods*. Thousand Oaks: Sage publications.

Minoiu, C., & Reddy, S. G. (2010). Development aid and economic growth: A positive long-run relation. *The Quarterly Review of Economics and Finance*, *50*(1), 27-39.

Mosley, P. (1980). Aid, savings and growth revisited. *Oxford Bulletin of Economics and Statistics*, *42*(2), 79-95.

Mosley, P. (1986). Aid-effectiveness: the micro-macro paradox. *IDS Bulletin*, *17*(2), 22-27.

Moss, T., Pettersson, G., & Van de Walle, N. (2006). *An aid-institutions paradox? A review essay on aid dependency and state building in sub-Saharan Africa* (Center for Global Development Working Paper No.74). Washington,D.C.: Center for Global Development.

Motala, S., & Pampallis, J.(2005). *Governance and finance in the South African schooling system: The first decade of democracy.* Johannesburg: CEPD and Wits EPU.

Motala, S., Dieltiens, V., & Sayed, Y. (2012). *Finding place and keeping pace: exploring meaningful and equitable learning in South African schools*. Cape Town: HSRC Press.

Motala, S., Morrow, S., & Sayed, Y. (2014). Gauteng Department of Education: A policy review In F. Maringe & M. Prew (Eds.) *Twenty years of education transformation in Gauteng 1994 to 2014: an independent review*. (Chapter 2). Johannesburg: Gauteng Department of Education.

Mouton, J., Wildschut, L., Richter, T. & Pocock, R. (2013). *Zenex Review Project* (Final Revised Report). Johannesburg: Evaluation Research Agency (ERA).

Moyo, D. (2009). *Dead aid: why aid is not working and how there is a better way for Afric*a. London: Allen Lane

Mphahlele, E. (2002). Education as community development–1990. *ES'KIA–Education African humanism & culture, social consciousness, literary appreciation*, 42-55.

Muller, J. (2001). Progressivism redux: ethos, policy, pathos. Introduction. In Kraak, A. & Young, M., (Eds.) *Education in Retrospect* (pp 59-72). Johannesburg: Ultra Litho (Pty) Limited.

Muller, J. (2004). Assessment, qualifications and the National Qualification Framework in South African schooling. In L. Chisholm (Ed.), *Changing class: Education and Social Change in post-apartheid South Africa* (pp. 221-245). Cape Town: HSRC Press.

Murnane, R. J., & Ganimian, A. J. (2014). *Improving educational outcomes in developing countries: lessons from rigorous impact evaluations*, NBER Working Paper. N. 20284. Cambridge: National Bureau of Economic Research.

Murphy, K.M., Shleifer, A., & Vishny, R.W. (1989). Industrialization and the Big Push. *The Journal of Political Economy*, *97*(5), 1003-1026.

Myburgh, O., Poggenpoel, M., & Van Rensburg, W. (2004). *Learners' experience of teaching and learning in a second or third language* (National Institute for Health and Clinical Excellence). Retrieved from http://www.nice.org.uk/

Naidoo, V. (2012). *Errors blamed for maths test 'disaster'*. Times Live. 9 December 2012.

Narsee, S. (2002). Culturally and linguistically diverse vocational training in South Africa. *Worlds of Learning: Globalisation and Multicultural Education*, 117-127.

National Education Policy Act (NEPA) No. 27 of 1996. Retrieved on 16 January 2016 from http://www.education.gov.za/Portals/0/Documents/Legislation/Acts/NATIONAL%20EDUCATION%20POLICY%20ACT,%20NO%2027%20OF%201996,%2028%2010%202011.pdf?ver=2015-01-30-102109-667

National Planning Commission (NPC) (2012). *National development plan vision 2030.* Pretoria: Presidency.

National Planning Commission (NPC). (2011). *Diagnostic overview*. Pretoria: Presidency.

National Treasury. (2009*). Adjusted estimates of national expenditure 2009.* Retrieved on 22 January 2016 from http://www.treasury.gov.za/documents/mtbps/2009/adjustments/National%20Treasury%20Ajusted%20Estimates.pdf

National Treasury. (2010*). Estimates of national expenditure 2010.* Retrieved on 22 January 2016 from http://www.treasury.gov.za/documents/national%20budget/2010/enebooklets/bookletvote34.pdf

National Treasury. (2011). *Estimates of national expenditure 2011*. Retrieved on 22 January 2016 from http://www.treasury.gov.za/documents/national%20budget/2011/ene/FullENE.pdf

National Treasury. (2014). *Budget speech*. Republic of South Africa, Pretoria: Government Printers.

National Treasury. (2015). *Estimates of national expenditure 2015*. Retrieved on 22 January 2016 from http://www.treasury.gov.za/documents/national%20budget/2015/ene/FullENE.pdf

Ndikumana, L. (2012). *Applying evaluation to development and aid: Can evaluation bridge the micro-macro gaps in aid effectiveness?* University of Massachusetts

Ndlovu, M. C. (2011). University-school partnerships for social justice in mathematics and science education: the case of the SMILES project at IMSTUS. *South African Journal of Education*, *31*(3), 419-433.

Ndlovu, M. C. (2011). Re-envisioning the scholarship of engagement: lessons from a university-school partnership project for mathematics and science teaching. *South African Journal of Higher Education*, *27*(7), 1397-1415.

NEPAD (New Partnership for Africa's Development) (2002), *Declaration on Democracy, Political, Economic and Corporate Governance*, signed in Durban 18 June 2002.

Newey, W. K., & Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrical*, *71*(5), 1565-1578.

Neuman, W. L. (2006). *Social research methods: qualitative and quantitative approaches* (6th ed.). Boston, MA: Allyn & Bacon.

Neumayer, E. (2003). The pattern of giving aid: the impact of good governance on development assistance. *Routledge Studies in Development Economics*, *34*, 394-411.

Nevondwe, L., & Matotoka, M. (2013). Promoting and protecting the right of access to basic education in South Africa. *The Thinker*, *5*(7), 8-13.

Ntuta, N. P., & Schurink, W. (2010). Towards sustaining performance in a Gauteng secondary school. *SA Journal of Human Resource Management*, *8*(1), 9-pages.

Nye, C., Schwartz, J., & Turner, H. (2006). Approaches to parent involvement for improving the academic performance of elementary school age children: a systematic review. *Campbell Systematic Reviews*, *2*(4).

Oakley, P. (1999). *The Danish NGO impact study: a review of Danish NGO activities in developing countries*. Oxford: INTRAC (The International NGO Training and Research Center).

O'Donoghue, T., & Punch, K. (2003). *Qualitative educational research in action: doing and reflecting.* London: Routledge.

Okun, A. (1975). *Equality and efficiency: the big trade-off*. Washington, DC.: The Brookings Institution

Olken, B. A. (2005). *Monitoring corruption: evidence from a field experiment in Indonesia* (No. w11753). Cambridge: National Bureau of Economic Research.

Ono, Y., & Ferreira, J. (2010). A case study of continuing teacher professional development through lesson study in South Africa. *South African Journal of Education*, 30(1), 59-74.

Organisation for Economic Co-operation and Development (OECD) (2008b). *Reviews of national policies for education*: *South Africa*. Paris: OECD.

Organisation for Economic Co-operation and Development (OECD) (2008a). *The Paris Declaration on Aid Effectiveness and the Accra Agenda for Action* (2005/2008). meParis: OECD.

Organisation for Economic Co-operation and Development (OECD) (2011). *Busan Partnership for Effective Development Co-operation*. Fourth High Level Forum on Aid Effectiveness, Busan, Republic of Korea, 29 November-1 December 2011. Retrieved from http://www.oecd.org/development/effectiveness/busanpartnership.htm

Organisation for Economic Co-operation and Development (OECD) (2012). *Policy framework for policy coherence for development* (Working Paper No. 1). Paris: OECD Office of the Secretary-General Unit for Policy Coherence for Development. Retrieved from http://www.oecd.org/development/pcd/50461952.pdf

Organisation for Economic Co-operation and Development (OECD) (2015). *OECD Data*. Paris: OECD. Retrieved on 26 November 2015 from https://data.oecd.org/searchresults/?r=%2Bf%2Ftype%2Findicators&r=%2Bf%2Ftopics_en%2Feducation

Organisation for Economic Co-operation and Development (OECD). (2011). *Final Country Evaluation Report* (Phase Two Evaluation of the Implementation of the Paris Declaration and Accra Agenda for Action in South Africa). Retrieved from https://www.oecd.org/derec/dacnetwork/47655903.pdf

Osterholm, M. T., Kelley, N. S., Sommer, A., & Belongia, E. A. (2012). Efficacy and effectiveness of influenza vaccines: a systematic review and meta-analysis. *The Lancet infectious diseases*, *12*(1), 36-44.

Owen, J. M., & Rogers, P. (1999). *Program evaluation: Forms and approaches*. Beverly Hills: Sage.

Owen, J.M. (2007). *Program evaluation: forms and approaches* (3rd ed.). New York: The Guilford Press.

Padayachee, P., Boshoff, H., Olivier, W., & Harding, A. (011). A blended learning Grade 12 intervention using DVD technology to enhance the teaching and learning of mathematics: original research. *Pythagoras*, 32(1), 1–8.

Pampallis, J. (2003). Education reform and school choice in South Africa. In D. Plank, & G. Sykes, *Choosing choice: School choice in international perspective*. New York: Teachers College Press.

Pampallis, J. (2004). The education business: private contractors in public education. *Changing class: education and social change in post-apartheid South Africa*, 421-433.

Papanek, G. F. (1972). The effect of aid and other resource transfers on savings and growth in less developed countries. *The Economic Journal*,*82*(327), 934-950.

Parker, D. (2006). Grade 10–12 mathematics curriculum reform in South Africa: A textual analysis of new national curriculum Statements. *African Journal of Research in Mathematics, Science and Technology Education*, *10*(2), 59-73.

Patton, M. Q. (2008). *Utilization-focused evaluation*. Beverly Hilss: Sage publications

Patton, M. Q. (2011). *Developmental evaluation: applying complexity concepts to enhance innovation and use*. New York: Guilford.

Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *British Medical Journal*, *2*, 1243-1246.

Peikes, D. N., Moreno, L., & Orzol, S. M. (2012). Propensity score matching. *The American Statistician*. Vol. 62, Issue 3.

Pereira, C., & Du Toit, R. (2012). Testing access: an assessment of epistemological access to numeracy.In Motala S, Dieltiens V & Sayed Y (eds) *Finding Place and Keeping Pace: Exploring Meaningful and Equitable Learning in South African Schools*. Cape Town: HSRC Press.

Perrin, B. (1998). Effective use and misuse of performance measurement. *American journal of Evaluation*, *19*(3), 367-379.

Perold & Associates (2012). *Business in support of basic education in South Africa*. Johannesburg: National Business Initiative and Sanlam.

Petrosino, A., Morgan, C., Fronius, T.A., Tanner-Smith, E.E., Boruch, R.F. (2012). Interventions in developing nations for improving primary and secondary school enrolment of children: A systematic review. *Campbell Systematic Reviews,* 2012 (19).

Picciotto, R. (2006). *Development effectiveness at the country level*. Bonn, Germany: German Development Institute.

Pillay, U., Tomlinson, R., & du Toit, J. (Eds.). Introduction. In U. Pillay, R. Tomlinson, & J. du Toit, *Democracy and delivery: Urban policy in South Africa*. Pretoria: HSRC Press.

Piper, B., Ralaingita, W., Akach, L., & King, S. (2016). Improving procedural and conceptual mathematics outcomes: evidence from a randomised controlled trial in Kenya. *Journal of Development Effectiveness*, 1-19.

Pitman, G. K., Feinstein, O. N., & Ingram, G. K. (Eds.). (2005). *Evaluating development effectiveness* (Vol. 7). New Brunswick, NJ: Transaction Publishers.

Prebisch, R. (1959). Commercial policy in the underdeveloped countries. *American Economic Review*, *49*(2), 251-273.

Prew, M. (2002). Towards a Model of Systemic School Improvement. *Draft Chapter of PhD Thesis*. Johannesburg: University of Witwatersrand

Prinsloo, C., & Kanjee, A. (2005). *Improving learning in South African schools: the Quality Learning Project (QLP) summative evaluation (200-2004).* Pretoria: Human Sciences Research Council.

Prinsloo, C. H. (2008). *Extra classes, extra marks? Report on the Plus Time Project*, Pretoria: HSRC

Public Finance Management Act (PFMA)1999. Legislation retrieved on 22 January 2016 from http://www.treasury.gov.za/legislation/pfma/default.aspx

Puhani, P. A. (2012). The treatment effect, the cross difference, and the interaction term in nonlinear "difference-in-differences" models. *Economics Letters*, *115*(1), 85-87.

Quigley, F. (2015). The Limits of Philanthropy. *Commonweal*, *142*(2), 9.

Ramalingam, B. (2011, January 31). Why the results agenda doesn't need results, and what to do about it [Web log post]. Retrieved from https://aidontheedge.info/2011/01/31/why-the-results-agenda-doesnt-need-results-and-what-to-do-about-it/

Rankin N., Schöer V., Sebastiao C., & Van Walbeek C. (2012). Predictors of academic performance: National Senior Certificate versus National Benchmark Test. *South African Journal of Higher Education*, 26(3),564-583.

Rasella, D., Aquino, R., Santos, C. A., Paes-Sousa, R., & Barreto, M. L. (2013). Effect of a conditional cash transfer programme on childhood mortality: a nationwide analysis of Brazilian municipalities. *The lancet*, *382*(9886), 57-64.

Ravallion, M. (2008). *Evaluation in the practice of development* (Policy Research Working Paper No. 4547). Washington, D.C.: World Bank.

Raudenbush, S.W. (1994). Random effects models. In H. Cooper, & L.V. Hedges (Eds.), *The handbook of research synthesis*. New York: Russell Sage Foundation.

Reddy, V. (2006). The state of mathematics and science education: Schools are not equal. In S. Buhlungu, J. Daniel, R. Southall, & J. Lutchman (Eds.) *State of the Nation: South Africa 2005-2006* (392-416). Cape Town: HSRC Press.

Reinikka, R., & Svensson, J. (2001). Confronting competition: investment, profit, and risk. *Uganda's Recovery: The Role of Farms, Firms, and Government*, 207-34.

Reinikka, R., & Svensson, J. (2006). Using micro-surveys to measure and explain corruption. *World Development*, *34*(2), 359-370.

Riddell, R. (2008). *Does foreign aid really work?*. Oxford: Oxford University Press.

Riddell, R. & Kruse, S.E. (1997). *Searching for impacts and methods: NGO evaluation synthesis study* (Report of Evaluation Study). Finland: Department for International Development Cooperation.

Riddell, R. C., Bebbington, A., & Peck, L. (1995). *Promoting development by proxy. An evaluation of the development impact of government support to Swedish NGOs*. Stockholm: SIDA.

Ritter, G., Albin, G., Barnett, J., Blankenship, V., & Denny, G. (2006). The effectiveness of volunteer tutoring programs: A systematic review. *Campbell Systematic Reviews*, *2*(7).

Robson, C. (2007). *How to do a research project: A guide for undergraduate students. Hoboken,* USA: Blackwell Publishing.

Roberts, J., & Schollar, E. (2006). *Meta-evaluation of interventions in math, science and language 1998-2006.* Johannesburg: ESA.

Roberts, J., & Schollar, E. (2011). *Meta-Evaluation of programmes and projects supported by the Zenex Foundation 2006-2011*. Johannesburg: ESA

Roche, C., & Kelly, L. (2005). Evaluating the performance of development agencies. In G. Pitman, O. Feistein, & G. Ingram, Vol.7. *Evaluating development effectiveness* (World Bank Series on Evaluation and Development). Edison, NJ: Transaction Publishers.

Roche, C. J. (1999). *Impact assessment for development agencies: Learning to value change*. Oxford: Oxfam.

Rogers, P.J. (2000). Causal models in program theory evaluation. In P.J. Rogers, A.J. Petrosino, T. Hacsi, and T.A. Huebner (Eds.), *Program theory evaluation: Challenges and opportunities* (New Directions in Evaluation Series, Issue 87, pp. 47-55). San Francisco, CA: Jossey-Bass.

Rogers, P.J. (2009). Matching impact evaluation design to the nature of the intervention and the purpose of the evaluation. *Journal of Development Effectiveness, 1*(3), 217-226.

Roithmayr, D. (2002). *The constitutionality of school fees in public education*. (Education Rights Project Issue Paper). Johannesburg: Wits Education Policy Unit, University of the Witwatersrand.

Roithmayr, D. (2003). Access, adequacy and equality: the constitutionality of school fee financing in public education. *South African Journal on Human Rights 19*(3), 397.

Romero, R. G., & Noble, M. (2008). Evaluating England's 'New Deal for Communities' programme using the difference-in-difference method. *Journal of Economic Geography*, 8(6), 759-778.

Rondinelli, D. A. (2013). *Development projects as policy experiments: An adaptive approach to development administration*. Routledge.

Roodman, D. (2008). *Through the looking-glass, and what OLS found there: on growth, foreign aid, and reverse causality* (Center for Global Development Working Papers 137). Washington, D.C.: Center for Global Development.

Rosenbaum, P.R., & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70 (1), 41-55.

Rosenberg, L. J., & Posner, L. D. (1979). The logical framework: a manager's guide to a scientific approach to design and evaluation. *Practical Concepts Incorporated, Washington, DC*.

Rosenstein-Rodan, P. (1943). Problems of industrialization of Eastern and Southeastern Europe. *Economic Journal*, *53* (210/211), 202-211.

Rosenstein-Rodan, P. (1961). Notes on the theory of the Big Push. In H.S. Ellis (Ed.), *Economic development for Latin America* (Proceedings of a conference held by the International Economic Association). London: Macmillan.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis*. New York: Russell Sage Foundation.

Rosenzweig, M. R., & Wolpin, K. I. (1986). Evaluating the effects of optimally distributed public programs: Child health and family planning interventions. *The American Economic Review*, *76*(3), 470-482.

Rossi, P.H., Lipsey, M.W., & Freeman, H.E. (2004). *Evaluation a systematic approach* (7th ed.). Thousand Oaks, CA: SAGE Publications, Inc.

Royse, D., Thyer, B. A., & Padgett, D. K. (2015). *Program evaluation: An introduction to an evidence-based approach*. Boston: Cengage Learning.

Ryan, E. D. (2007). *The value of school-initiated professional development in South African schools: A Case Study of Two Schools in Two Gauteng Districts*. Doctoral dissertation. Johannesburg: University of the Witwatersrand.

Sachs, J. (2005). *The end of poverty: how we can make it happen in our lifetime*. London: Penguin Books.

Salanti, G. (2012). Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Research Synthesis Methods*, *3*(2), 80-97.

Sale, J.E.M., Lohfeld, L.H., & Brazil, K. (2002). Revisiting the quantitative-qualitative debate: Implications for mixed-methods research. *Quality and Quantity*, *36* (1), 43-53.

Salter, K. L., & Kothari, A. (2014). Using realist evaluation to open the black box of knowledge translation: a state-of-the-art review. *Implementation science*, *9*(1), 1.

Sanson-Fisher, R. W. Bonevski, B., Green, L. W., & D'Este, C. (2007). Limitations of the randomized controlled trial in evaluating population-based health interventions. *American Journal of Preventive Medicine*, 33(2), 155-161.

Saretsky, G. (1972). The OEO PC Experiment and the John Henry Effect. *The Phi Delta Kappan*, 53(9), 579-581.

Sayed, Y., & Motala, S. (2012). Getting in and staying there: exclusion and inclusion in South African schools. *Southern African Review of Education*, 18(2), 105-118.

Sayed, Y., Kanjee, A., & Nkomo, M. (2013). *The search for quality education in post-apartheid South Africa: Interventions to improve learning and teaching*. Cape Town: HSRC Press.

Schaffer, A. & Watters, K. (2010). *Barriers and bridges to learner understanding and performance in Grade 11, English, Maths and Science* (Report on a study at six Dinaledi schools for the Zenex Foundation). Johannesburg: Zenex Foundation.

Schafft, K. A., & Greenwood, D. J. (2003). Promises and dilemmas of participation: action research, search conference methodology, and community development. *Journal of Community Development Society*, 34(1), 18-35.

Schiavo-Campo, S. (1999). "Performance" in the public sector. *Asian Journal of Political Science*, *7*(2), 75-87.

Schmidt, F. L., & Hunter, J. E. (2014). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks: Sage publications.

Schoer, V., Ntuli, M., Rankin, N., Sebastiao, C., & Hunt, K. (2010). A blurred signal? The usefulness of National

Schollar, E. (2013). The Learning for Living Project: a book-based approach to learning of language in South African primary schools. In Y. Sayed, A. Kanjee, & M. Nkomo (Eds.), *The search for quality education in post-apartheid South Africa* (pp. 237-256). Cape Town, South Africa: HSRC Press.

Senior Certificate (NSC) Mathematics marks as predictors of academic performance at university level. *Perspectives in Education*, 28(2), 9–18.

Schollar, E. (1999). *Evaluation of the Amplats School Development Programme*. Johannesburg: Eric Schollar and Associates

Schollar, E. (2001). A Review of Two Evaluations of the application of The READ Primary Schools Programme in the Eastern Cape province of South Africa. *The International Journal of Educational Research (IJER),* 35(2).

Schollar, E (2002). *The Evaluation of the Pilot EQUIP project of the National Business Initiative* (Final Report). Johannesburg: ESA

Schollar et al. (2005). *The evaluation of the Learning for Living Project*. Johannesburg: Eric Schollar and Associates.

Schollar, E. (2007). *Towards evidence based education development in South Africa.* Johannesburg: Eric Schollar and Associates.

Schollar, E. (2008). *The Primary Mathematics Research Project 2004-2007: towards evidence-based educational development in South Africa* (Final Report). Johannesburg: Eric Schollar & Associates.

Schollar, E. (2015). *The Primary Mathematics Research Project: 2004-2012. An evidence based programme of research into understanding and improving the outcomes of mathematical education in South African primary schools.* Unpublished PhD Thesis. Cape Town: University of Cape Town.

Schollar, E., & Mouton, J. (2014). *Epoch and Optima: Public Schools Math Challenge Programme. An impact evaluation* (Final Report). Johannesburg: ESA & ERA.

Schulz, N.-S. (2010). *Triangular cooperation in the context of aid effectiveness- Experiences and views of EU donors*. Madrid: Spanish Agency for International development cooperation.

Schwab, K. (Ed.). (2011). *The Global Competitiveness Report*. Geneva, Switzerland: World Economic Forum (WEF).

Scriven, M. (1976) Maximizing the power of causal investigations: the modus operandi method. *GV Glass Evaluation Studies Review Annual*. Beverly Hills. SAGE

Scriven, M. (1991). *Evaluation thesaurus*. New Castle: Sage Publications

Scriven, M. (1994). The fine line between evaluation and explanation. *American Journal of Evaluation, 15*(1), 75-77.

Selener, J.D. (1997). *Participatory action research and social change: approaches and critique*. New York: Cornell University.

Sen, A. (1999). *Development as freedom*. New York, NY: Oxford University Press.

Shadish, W.R, Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Wadsworth: Cengage Learning.

Shadish, W. R., & Cook, T. D. (2009). The renaissance of field experimentation in evaluating interventions. *Annual Review of Psychology*, *60*, 607-629.

Shively, G., & Galopin, M. (2013). *Overview of benefit-cost analysis*. West Lafayette, IN: Purdue University. Retrieved from http://www.agecon.purdue.edu/staff/shively/COURSES/AGEC406/reviews/bca.htm.

Sibanda, L. (2013). *The linguistic challenges of the grade 4 mathematics ANAs and the way in which teachers manage these demands in their preparation for the ANAs*. Unpublished PhD Research Proposal, Rhodes University, Education Department, Grahamstown.

Sidiropoulos, E., Pérez Pineda,J.A., Chaturvedi, S., & Fues, T. (Eds.) (2015). *Institutional architecture and development: responses from emerging powers*. Johannesburg: South African Institue of International Affairs (SAIIA).

Simkins, C. (2010). *The maths and science performance of South Africa's public schools: some lessons from the past decade* (Executive Summary No. 1). Johannesburg: CDE.

Simkins, C., & Paterson, A. (2005). *Learner performance in South Africa: social and economic determinants of success in language and mathematics*. HSRC Press. Pretoria.

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32(9), 752.

Sobhy, H. (2012). The de-facto privatization of secondary education in Egypt: A study of private tutoring in technical and general schools. *Compare: A Journal of Comparative and International Education*, *42*(1), 47-67.

South African Institute of Race Relations (SAIRR) (2011). South African Institute of Race Relations South Africa Survey 2010-2011. Accessed on 22 February 2016 from http://www.sairr.org.za/services/publications/south-africa-survey/south-africa-survey-online-2010-2011/education.[Webcitation]

South African Schools Act (SASA), 1996 (Act No. 84 of 1996). Retrieved from http://www.acts.co.za/south-african-schools-act-1996/index.html01/05/2016[Style

Spady, W. (1994). Choosing outcomes of significance. *Educational Leadership 51* (5), 18–23.

Spady, W. G. (2008). It's time to end the decade of confusion about OBE in South Africa. *Suid-Afrikaanse Tydskrif vir Natuurwetenskap en Tegnologie* 27 (1): 17-29.

Spaull, N. (2011). *A preliminary analysis of SACMEQ III South Africa* (Stellenbosch Economic Working Papers 11/11). Stellenbosch: University of Stellenbosch.

Spaull, N. (2013). *South Africa's education crisis: the quality of education in South Africa 1994-2011* (CDE Report). Johannesburg: Center for Development and Enterprise.

Spaull, N. (2015). *My opening remarks at the OR Tambo Debate and an afterword…* Retrieved from 27 November 2015 from https://nicspaull.com/2015/07/.

Schreiner, M. (2003). A cost-effectiveness analysis of the Grameen Bank of Bangladesh. *Development Policy Review*, *21*(3), 357-382.

Stand, J. (2000). The" Hawthorne Effect"-what did the original Hawthorne studies actually show? *Scand J Work Environ Health*, *26*(4), 363-367.

Stilwell, L.C., & Minnitt, R.C.A (2006). *Is platinum paying its rent?* Presented at the International Platinum Conference 'Platinum Surges Ahead'. Johannesburg: The Southern African Institute of Mining and Metallurgy.

Strauss, J., & Burger, M. (2000), *Monitoring learning achievement project.* Pretoria. Department of Education,

Sturtz, S., & Bender, R. (2012). Unsolved issues of mixed treatment comparison meta-analysis: Network size and inconsistency. *Research Synthesis Methods*, *3*(4), 300-311.

Suchman, E. A. (1967). *Principles and practice of evaluative research. An Introduction to Social Research.* New York: Appleton-Century-Crofts, 327-351

Sujatha, K., & Rani, P. G. (2011). *Management of secondary education in India*. New Delhi: Shipra and National University of Educational Planning and Administration (NUEPA).

Sumner, A. (2012). *Global poverty and the "New Bottom Billion" revisited: exploring the paradox that most of the world's extreme poor no longer live in the world's poorest countries* (Working Paper). Retrieved from https://www.ids.ac.uk/files/dmfile/AndySumnerGlobalPoverty14May2012.pdf

Summers, L. H. (1994). *Investing in all the people: educating women in developing countries.* Washington, DC: World Bank.

Summers, H.C. (2008). *Maths Centre/Zenex Foundation Project: upgrading of Teacher Qualifications at Foundation Phase KwaZulu-Natal and the Eastern Cape (2005-2008)* (Summative Evaluation Report). Johannesburg: Zenex.

Sutton, A.J., Jones, D.R., Abrams, K.R., Sheldon, T.A., & Song, F. (2000). *Methods for meta-analysis in medical research*. London: John Wiley

Svensson, U. P., Fred, R. I., & Vanderkooy, J. (1999). An analytic secondary source model of edge diffraction impulse responses. *The Journal of the Acoustical Society of America*, *106*(5), 2331-2344.

Swilling, M., & Russel, B. (2002). *The size and scope of the non-profit sector in South Africa* (Monograph). Johannesburg: Graduate School of Public and Development Management (P&DM), University of the Witwatersrand and The Centre for Civil Society, University of Natal.

Tanzi, V. (2016). Public Finance in Developing Countries: An Introduction. *Handbook of Research on Public Finance in Europe and the MENA Region*, 1.

Taylor, N., & Vinjevold, P. (1999). *Getting learning right: report of the President's education initiative research project*. Johannesburg: JET.

Tan, J., Lane, J., & Lassibille, G. (1999). Student outcomes in Philippine Elementary Schools: An evaluation of four experiments. *World Bank Economic Review, 13*(3), 493-502.

Taylor, E. W. (2007). An update of transformative learning theory: a critical review of the empirical research (1999–2005). *International journal of lifelong education*, *26*(2), 173-191.

Taylor, N., & Prinsloo, C. (2005). *The quality learning project lessons for high school Improvement in South Africa.* Johannesburg: Joint Education Trust.

Taylor, N., Fleisch, B., & Shindler, J. (2008). *Changes in education since 1994.* Pretoria: Presidency.

Taylor, N. (2009). Standards-based accountability in South Africa. *School Effectiveness and School Improvement*, *20* (3), 341-356.

Taylor, N., Gamble, J., Spies, M. & Garisch, C.(2012). School leadership and management. In N. Taylor, S. van der Berg, & T. Mabogoane (Eds.), *What makes schools effective*? (Report of South Africa's National School Effectiveness Study). Cape Town: Pearson Education.

Taylor, N., Muller, J., & Vinjevold, P. (2003). *Getting schools working*. Cape Town: Pearson Education South Africa.

Taylor N, (2011) *A five-year plan for South African schooling: a report commissioned by the National Planning Commission*. Johannesburg: JET Education Services

Taylor, N., van der Berg, S., & Mabogoane, T. (2013). *What makes schools effective?* (Report of the National Schools Effectiveness Study). Cape Town: Pearson Education South Africa.

Taylor, S. (2011). *Uncovering indicators of effective school management in South Africa using the National School Effectiveness Study*. Stellenbosch: Stellenbosch University.

Taylor, S. Watson, P. (2013). *The impact of study guides on matric performance: evidence from a randomised experiment*. Stellenbosch Economic Working Paper. University of Stellenbosch.

Taylor, S., & Coetzee, M. (2013). *Estimating the impact of language of instruction in South African primary schools: a fixed effects approach*. Stellenbosch Economic Working Paper. Stellenbosch University.

Taylor, S., & Yu, D. (2009). Socio-economic status and educational achievement: Does education provide a stepping stone out of poverty in South Africa? Stellenbosch Economic Working Paper. Stellenbosch University.

The Multilateral Organisation Performance Assessment Network (MOPAN) (2011). *MOPAN Annual Report 2011*. Retrieved from www.mopanonline.org/es/.../MOPAN_Annual_Report_2011_FINAL.pd

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica: journal of the Econometric Society*, 24-36.

Todd, P. (2012). *Non-experimental and quasi-experimental methods for impact evaluation,* Lecture at the CLEAR/3ie Impact evaluation technical course in Johannesburg, September 2012. Johannesburg: University of Witwatersrand.

Torjman, S., & Makhoul, A. (2012). *Community-led development*. Ottawa, ONT: Caledon Institute of Social Policy.

Torvatn, H. (1998). Evaluation of industrial development in Norway. *New Directions for Evaluation*, *1998*(77), 71-87.

Treweek, S., & Zwarenstein, M. (2009). Making trials matter: pragmatic and explanatory trials and the problem of applicability. *Trials*, *10*(1), 1.

Trialogue (2005). *The CSI handbook* (8th ed.). Cape Town: Trialogue.

Trialogue. (2010). *Trialogue 2010 Annual Report*. Cape Town: Trialogue. Retrieved from www.trialogue.co.za

Trialogue. (2011). *Trialogue 2011 Annual Report*. Cape Town: Trialogue. Retrieved from www.trialogue.co.za

Trialogue. (2013). *Trialogue 2013 Annual Report*. Cape Town: Trialogue. Retrieved from www.trialogue.co.za

Trialogue. (2015). *The Trialogue 2015 CSI handbook*. Cape Town: Trialogue. Retrieved from www.trialogue.co.za

Trochim, W. M., & Donnelly, J. P. (2001). *Research methods knowledge base*. New York: Cornell University

Trochim, W., & Donnelly, J.P. (2006). *The research methods knowledge base* (3rd ed.). Boston, MA: Cengage Learning.

Trochim, W., & Land, D. (1982). Designing designs for research. *The Researcher*, *1*(1), 1-6.

Udry, C. (1996). Gender, agricultural production, and the theory of the household. *Journal of Political Economy*, *104*(5),1010-1046.

Umalusi (2004) *Have the standards of the Senior Certificate Examination declined? Summary report on the evaluation of the senior certificate examination*. Pretoria: Umalusi.

United Nations (UN) (2000). *United Nations Millennium Declaration*. Retrieved on 13/06/2011 from http://www.un.org/millennium/declaration/ares552e.htm

United Nations (UN) (2002). United Nations Millennium Development Goals. Retrieved on 13/06/2011 from http://www.un.org/millenniumgoals/

United Nations (UN) (2002). *We can end poverty: millennium development goals and beyond 2015*. Retrieved on 13/06/2011 from http://www.un.org/millenniumgoals/bkgd.shtml

United Nations (UN) (2003). *Monterrey consensus on financing for development*. Retrieved on 13/06/201 from http://www.un.org/esa/ffd/monterrey/

United Nations (UN) (2013). *A new global partnership: eradicate poverty and transform economies through sustainable development* (Report of the High-Level Panel of Eminent Persons on the post-2015 development agenda). New York: United Nations.

United Nations Department of Economic and Social Affairs (UNDESA) (2010). *Development cooperation for the MDGs – Maximizing Results* (International Development Cooperation Report). New York: UNDESA.

United Nations Development Group (UNDG) (2008). *Third high level forum on aid effectiveness*, Accra, Ghana, 2-4 September 2008 (UNDG Report). New York: UNDG.

United Nations Development Group (UNDG) (2011). *Results-based management guidance and policy*. New York: UNDG

United Nations Development Programme (UNDP) (2002). *Handbook on monitoring and evaluating for results.* New York: UNDP

United Nations Development Programme (UNDP) (2015). *Goal 2: Achieve universal primary education.* Retrieved from http://www.un.org/millenniumgoals/education.shtml

UNESCO. (2000). *World Education Forum: final report*. Paris: UNESCO

UNESCO (2015). *Rethinking education: towards a common global good?* Paris, France: UNESCO Publishing.

UNICEF (2011), *Tracking public expenditure and assessing service quality in early childhood development in South Africa*, Pretoria: UNICEF, DBE & DSD

Van der Klaauw, W. (1997). *A regression-discontinuity evaluation of the effect of financial aid offers on college enrollment* (Working Paper 97-10). NY: C.V. Starr Center for Applied Economics, New York University.

Van der Berg, S., & Louw, M. (2007). Lessons learnt from SACMEQII: South African student performance in regional context. *University of Stellenbosch, Department of Economics and Bureau for Economic Research Working Paper*, *16*(07).

Van der Berg, S. (2008). How effective are poor schools? Poverty and educational outcomes in South Africa. *Studies in Educational Evaluation*, *34*(3), 145–154.

Van der Berg, S. (2011, 08 April). SA education: The poorest choice. Why do South African children perform worse than those in poorer countries? *Mail and Guardian*. Accessed http://mg.co.za/article/2011-04-08-sa-education-the-poorest-choice/

Van der Berg, S. & Spaull, N. (2012, 07 Dec). Vast improvements in pupils' national test results not possible. Interview with *Mail and Guardian*. Accessed at mg.co.za/ANAs2012

Van Houwelingen, H. C., Arends, L. R., & Stijnen, T. (2002). Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in medicine*, *21*(4), 589-624.

Van Staden, S., & Bosker, R. (2014). Factors that affect South African reading literacy achievement: evidence from prePIRLS 2011. *South African Journal of Education*, *34*(3), 01-09.

Veriava, F., & Wilson, S. (2005). A critique of the proposed amendments on school funding and school fees. *ESR Review, 6*(3), 9.

Vilakazi, N (2015). *Black African learner performance: an in-depth study.* Presentation made at Department of Basic Education Research Seminar 29th July 2015, Pretoria

Vorster, C., Mayet, A., & Taylor, S. (2013). *The language of teaching and learning in South Africa schools.* University of Stellenbosch, Bureau for Economic Research.

Wadvalla, I. (2005). *The implementation of the IQMS in two different schools in Gauteng.* Unpublished BEd Hons project. Johannesburg: University of the Witwatersrand.

Walker, E., Hernandez, A.V., & Kattan, M.W. (2008). Meta-analysis: its strengths and limitations. *Cleveland Clinic Journal of Medicine*. 75(6), 431-439.

Watson, D., Broemeling, A. M., Reid, R. J., & Black, C. (2014). *A results-based logic model for primary health care: laying an evidence-based foundation to guide performance measurement, monitoring and evaluation.* Vancouver: University of British Columbia Library

Weisner, E. (2005). The 'resource restriction condition' for evaluation effectiveness. In G.K. Pitman, O.N. Feinstein, & G.K. Ingram (Eds.), *Evaluating development effectiveness* (pp. 337-346) (World Bank Series on Evaluation and Development, Vol.7). Piscataway, NJ: Transaction Publishers.

Weiss, C. H. (1997). Theory-based evaluation: past, present, and future. *New directions for evaluation*, *1997*(76), 41-55.

Weiss, C. (1998). *Evaluation: methods for studying programs and policies*. Englewood Cliffs, NJ: Prentice Hall.

Weisskopf, T. E. (1972). The impact of foreign capital inflow on domestic savings in underdeveloped countries. *Journal of international Economics*,*2*(1), 25-38.

White, H. (2005). *Challenges in evaluating development effectiveness* (Working Paper No. 242). Brighton: IDS.

White, H. (2008). Of probits and participation: the use of mixed methods in quantitative impact evaluation. *IDS Bulletin*, *39*(1), 98–109.

White, H. (2009a). *Some reflections on current debates in impact evaluation* (3ie Working Paper No. 1). New Delhi: International Initiative for Impact Evaluation.

White, H. (2009b). Theory based impact evaluation: principles and practice. *Journal of Development Effectiveness*, *1*(3), 271-284.

White, H., & Phillips, D. (2012). Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework. *International Initiative for Impact Evaluation, New Delhi.*

White, H. (2012). *Theory based impact evaluation: principles and practice*. Presentation at the CLEAR impact evaluation course, March 2012, Shanghai, China.

White, H. (2013). The use of mixed methods in randomized control trials. In D.M. Mertens, & S. Hesse-Biber (Eds.), *Mixed methods and credibility of evidence in evaluation*. San Francisco, CA: Jossey-Bass.

White, H., & Phillips, D. (2012). *Addressing attribution of cause and effect in small in impact evaluations: towards an integrated framework.* New Delhi: International Initiative for Impact Evaluation.

Wills, G (2015). *Informing principal policy reforms in South Africa through data-based evidence.* Presentation made at Department of Basic Education Research Seminar 29th July 2015, Pretoria

Wilson, S.J., Tanner-Smith, E.E., Lipsey, M.W., Steinka-Fry, & Morrison, J. (2011). Dropout prevention and intervention programs: Effects on school completion and dropout among school-aged children and youth. *Campbell Systematic Reviews* 2011 (8)

Wolfensohn, J.D. (2005). *Voice for the world's poor: selected speeches and writings of World Bank President James D. Wolfensohn, 1995-2005*.Washington, D.C.: World Bank.

Wolhuter, C.C. (2006). Teacher training in South Africa: past, present and future. *Education Research and Perspectives*, *33*(2), 124-139.

Woods, N. (2008). Whose aid? Whose influence? China, emerging donors and the silent revolution in development assistance. *International Affairs*,*84*(6), 1205-1221.

Woolman, S., & Fleisch, B. (2006). South Africa's unintended experiment in school choice: how the National Education Policy Act and the Employment of Educators Act create the enabling conditions for quasi-markets. *Education and the Law*, *18*(1), 31–75.

Woolman, S., & Fleisch, B. (2008). Democracy, social capital and school governing bodies in South Africa. *Education and the Law*, *20*(1): 47.

Woolman, S., & Fleisch, B. (2009). *The constitution in the classroom: law and education in South Africa 1994-2008.* Pretoria: Pretoria University Law Press.

World Bank. (2015a). *Incheon Declaration: education 2030 - towards inclusive and equitable quality education and lifelong learning for all*. Washington, D.C.: World Bank Group. Retrieved from http://documents.worldbank.org/curated/en/2015/12/25662441/incheon-declaration-education-2030-towards-inclusive-equitable-quality-education-lifelong-learning-all

World Bank (2015b). *Annual report 2015*. Washington, D.C.: World Bank.

World Bank (2016). *GINI index (World Bank estimate)*. Retrieved on 13 February 2016 from http://data.worldbank.org/indicator/SI.POV.GINI?page=1

Zief, S. G., Lauver, S., & Maynard, R. A. (2006). Impacts of after-school programs on student outcomes: A systematic review. *Campbell Systematic Reviews*, *2* (3).

Zoccai, G. B. (2014). *Network meta-analysis*. Rome: Nova Science Publishers, Inc.