What works and what doesn't

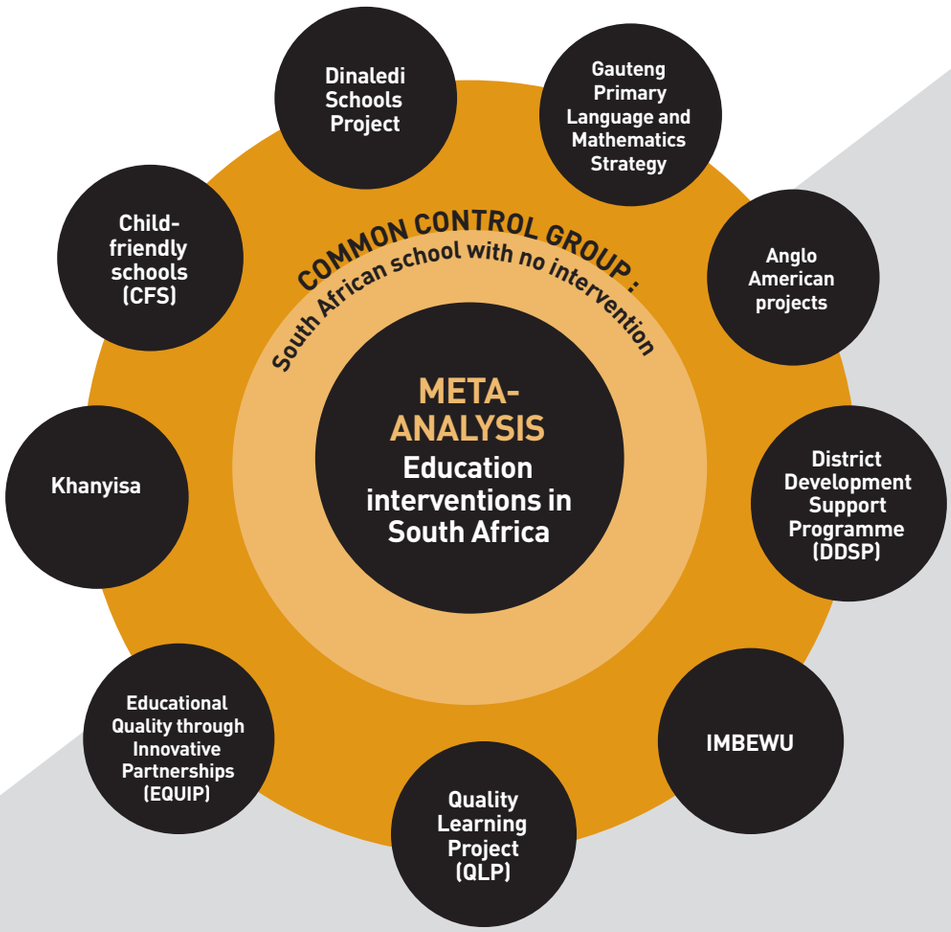
A meta-analysis of South African education interventions

META-ANALYSIS:

A statistical technique that pools or compares results from different evaluations.

In the past two decades various public and private institutions have implemented programmes and initiatives to address the education challenges in South Africa, but reliable evidence on what is truly working is very limited.

This aim of this meta-analysis is to explore the various contextual factors and design features that influence the magnitude of the effects reported in education impact studies. The meta-analysis investigates investors and policy-makers with a synthesis of 15 years of learning, in order **to inform new programming, and improve efforts to address education challenges**.



clear
Centers for Learning on Evaluation and Results
ANGLOPHONE AFRICA

ZENEX FOUNDATION

SOUTH AFRICAN SCIENCE EDUCATION INTERVENTIONS (2003–2015)

The forest plot below summarizes the effectiveness of science education interventions in South African high schools. The table includes each intervention's effect size (measured in Hedge's g) and the precision with which each effect size was measured (standard error). Effect size and precision are graphically depicted in the final column. The box represents the effect size while the horizontal line (confidence interval) measures the precision. The rule is: The bigger the box, and the shorter the confidence interval (horizontal line), the more precise the estimation.

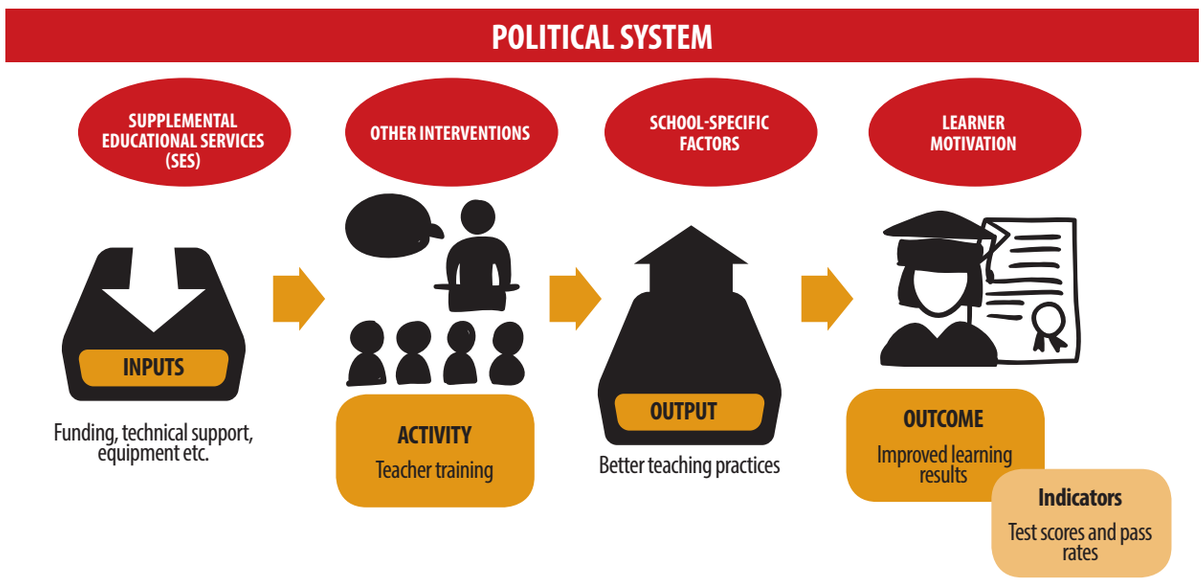
Intervention	Study	Subject	Grade	Hedge's g	Standard error	Hedge's g and 95% confidence interval (CI)
Dinaledi C	World Bank DIME (2010)	Science SG	12 - Sec	0.550	0.077	
Dinaledi D	World Bank DIME (2010)	Science HG	12 - Sec	0.269	0.076	
Problem-based learning	Loggerenberg-Hattingh (2003)	Science	10 - Sec	0.227	0.169	
Radical M&S Winter School F	Besharati (2014)	NSC Pass	12 - Sec	0.211	0.225	
Radical M&S Winter School C	Besharati (2014)	Science SG	12 - Sec	0.181	0.225	
MTG Study Guides G	Taylor & Watson (2015)	Geography	12 - Sec	0.141	0.019	
MTG Study Guides D	Taylor & Watson (2015)	Life sciences	12 - Sec	0.106	0.018	
Radical M&S Winter School E	Besharati (2014)	NSC Bachelor	12 - Sec	0.079	0.224	
Amplats E	Besharati (2014)	NSC Pass	12 - Sec	0.059	0.132	
Amplats D	Besharati (2014)	NSC Bachelor	12 - Sec	0.032	0.132	
Amplats G	Besharati (2014)	Science SG	12 - Sec	0.012	0.132	
MTG Study Guides C	Taylor & Watson (2015)	Economic science	12 - Sec	-0.036	0.024	
Mine Proximity F	Besharati (2014)	Science HG	12 - Sec	-0.114	0.078	
Mine Proximity E	Besharati (2014)	Science SG	12 - Sec	-0.154	0.078	
Mine Proximity H	Besharati (2014)	NSC Bachelor	12 - Sec	-0.160	0.078	
Mine Proximity G	Besharati (2014)	NSC Pass	12 - Sec	-0.168	0.078	
Amplats F	Besharati (2014)	Science HG	12 - Sec	-0.173	0.132	
Radical M&S Winter School D	Besharati (2014)	Science HG	12 - Sec	-0.254	0.225	
AVERAGE				**0.045**	**0.039**	

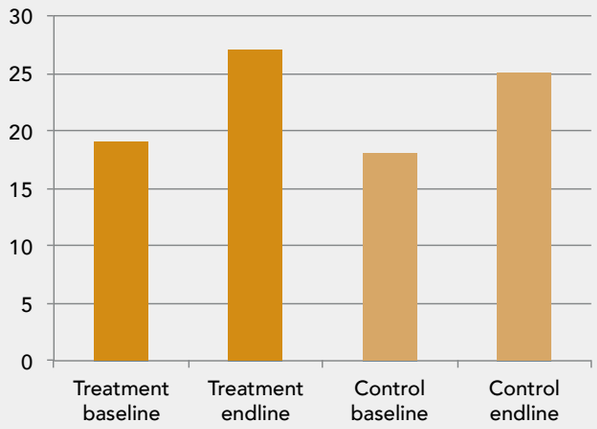
A THEORY OF CHANGE

The diagram below illustrates a typical school intervention theory of change (ToC). If implemented correctly, the teacher training intervention is expected to bring about better teacher practices in the classroom. Improved teaching is expected to result in better learning outcomes.

The results of the ToC may be affected by the confounding variables (in red).

Impact evaluations measure the progress of outcomes. In this ToC the outcomes are learning results which are measured by the test-score and pass-rate indicators. Indicators should only include publicly recognized tests or tests designed by evaluators.





NOTE: 95% confidence intervals were indicated

RCUP Pinewood study: The need for a comparator

The evaluation of the RCUP Pinewood catch-up programme showed that there was improvement in the treatment group, but, because the control group showed the same level of improvement, it was concluded that this was not necessarily as a result of the intervention.

To evaluate the intervention accurately, the team had to compare the magnitude of the differences in both groups.

Case study (often qualitative)	
Before-and-after assessment	
Programme and control group comparison	
Difference-in-difference	

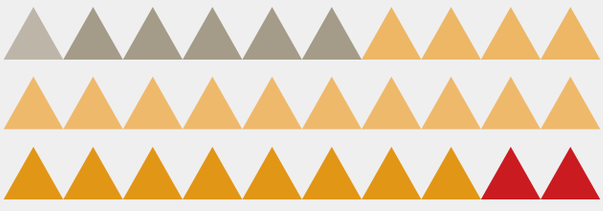
Types of education evaluation

Different evaluation methods have varying levels of statistical confidence. In the table on the previous page, the level of reliability improves as one goes down the list. A study design that includes a before-and-after assessment ***as well as a counterfactual*** is regarded as the most reliable.

COUNTERFACTUAL:

What would have happened in the absence of an intervention.

RISK OF BIAS (NUMBER OF EVALUATIONS)



Minimal (3%) Low (17%) Moderate (47%)
High (27%) Critical (7%)

AVERAGE EFFECT SIZE

Academic Government Private

Most education impact evaluations are problematic

Evaluation studies are subject to numerous threats to validity and risk of biases that may render findings untrustworthy. Although there are many types of risk-of-evaluation bias for this study (and many like it), the five most important are **selection bias, implementation bias, data and testing bias, evaluator bias and motivational bias**.

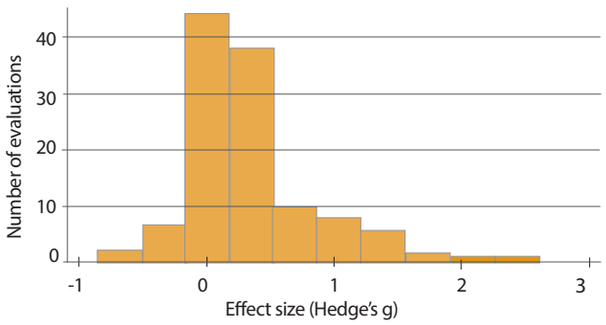
Risk-of-evaluation bias can be mitigated if selection bias has been appropriately addressed and programme implementation and causality are carefully controlled. Any bias, assumptions, errors or problems in any study must be carefully examined and reported in a meta-analysis.

During this review we found that the majority of education impact evaluations carry a moderate-to-high risk of bias.

FINDINGS OF THE META-ANALYSIS

1. Few interventions have had a substantial impact on learning outcomes

The histogram frequency plot below shows that although many of the included studies reported a high effect size, an equally significant number reported a negative effect size.



2. Simple interventions are often more effective

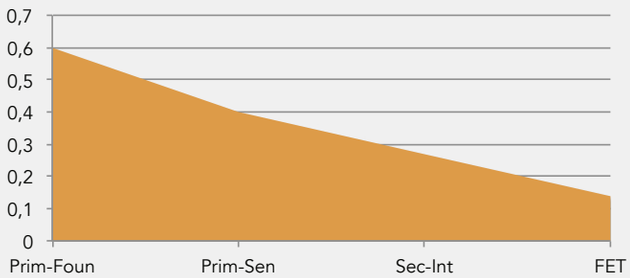
The study found that if simple interventions like providing learning materials are managed and designed well they may be more effective than expensive and more complex designs. Providing learning and teaching support materials (LTSM) can sometimes be more effective than complex and expensive whole-school development interventions.

EFFECT SIZE: COMPARISON OF INTERVENTIONS

Intervention type	Average effect size	Number of interventions
LTSM	0.593	36
Teacher-focussed	0.267	24
Whole-school intervention	0.204	38
Learner-focussed	0.158	21

3. Schooling phase and grade of assessment are the strongest predictors of impact

Interventions implemented among younger learners had a larger effect size, which is consistent with broader international work on child development and education. The figure shows the relationship between school phase and effect size. In most cases effect size decreases as the school phase increases. This suggests that in drafting policy there should be greater investment in early-phase interventions.



4. Context and study methodology influence effect size

We cannot take results of impact studies at face value – context and study methodology have a strong influence on effective size and cost effectiveness return. Factors such as assessment tools, evaluation method, units of analysis, and timing of evaluation all affect the results.

In evaluating the Back to the Basics (B2B) workbooks, the two evaluators (Fleisch, Taylor et al, 2010, and Schollar, 2015) had **opposite results**. Even though both evaluations were high quality randomised control trials (RCT), methodological problems and limitations produced different results.

EVALUATION OF BACK TO BASICS (B2B) WORKBOOKS

	Eric Schollar & Associates	Wits University and Joint Education Trust (JET)
Effect size	+1.34	-0.17
Programme dosage	high	high
Evaluation design	RCT	RCT
Grade	6	6
Location	Limpopo	Gauteng
Test instrument	ESA	JET
Evaluators	Internal	External
Counterfactual	No workbooks	Textbook (ESP)

5. Cost effectiveness analysis (CEA) is useful but difficult to do properly

CEA is a complex exercise that requires time, political will and technical rigor. Some of the challenges encountered when attempting to conduct a cost-benefit analysis of the various interventions are:

- Political and commercial sensitivities prevent institutions from disclosing financial information to external parties and where it is made available it is often unclear, inaccurate, incomplete or inconsistent across the various organizations/programmes
- Much of the cost information is estimated, and expressed as budgetary commitments rather than actual spend
- Some interventions are made possible by in-kind donations (free goods or voluntary services) that are often not reflected in budgets

CONCLUSIONS

1. Few interventions have had a substantial impact on learning outcomes in the South African public school system.
2. Although no authoritative conclusions on the best interventions have been reached, there is some indication on when, where, how and why interventions present with a bigger impact.
3. Context and study methodology – assessment tools, evaluation method, units of analysis and timing of evaluation – all influence effect size.
4. Schooling phase and grade of assessment are the strongest predictors of high impact.
5. Well-designed, well-delivered LTSMs are cost-effective and may have more impact than complex, expensive, whole-school development programmes.
6. The high effect sizes that result from academic experiments might not be repeated when the experiments are replicated, scaled or taken to policy spaces.
7. Cost effectiveness analysis is useful for policymakers, but is difficult to do properly. It is a complex exercise that requires time, political will and technical rigor.