RESEARCH ON SOCIO-ECONOMIC POLICY (ReSEP)
STELLENBOSCH UNIVERSITY

CONFERENCE

QUANTITATIVE APPLICATIONS IN EDUCATIONAL RESEARCH

Presentation

Evidence-Based Educational Development: The Primary Mathematics Research Project (2004-2012)

Eric Schollar

September 2016

Educational and evaluation programme research has not learned a great deal about improving learner performance across the system.

- What works?
- Why does it work?
- Under what conditions does it work?
- Under what conditions can it be expected to continue to work in replication at large scales within the routine school system?

What to do about education research's credibility gaps?

Become more scientific.

Increasing use of experimental and quasi-experimental designs for causal inference.

Developing collaboration between sociologists, educationalists and econometricians is yielding more reliable and useful results.

Models of independent variables which are most closely associated with higher levels of learner performance.

BUT we are not all the way there yet and nor is it as "simple" as it once seemed!

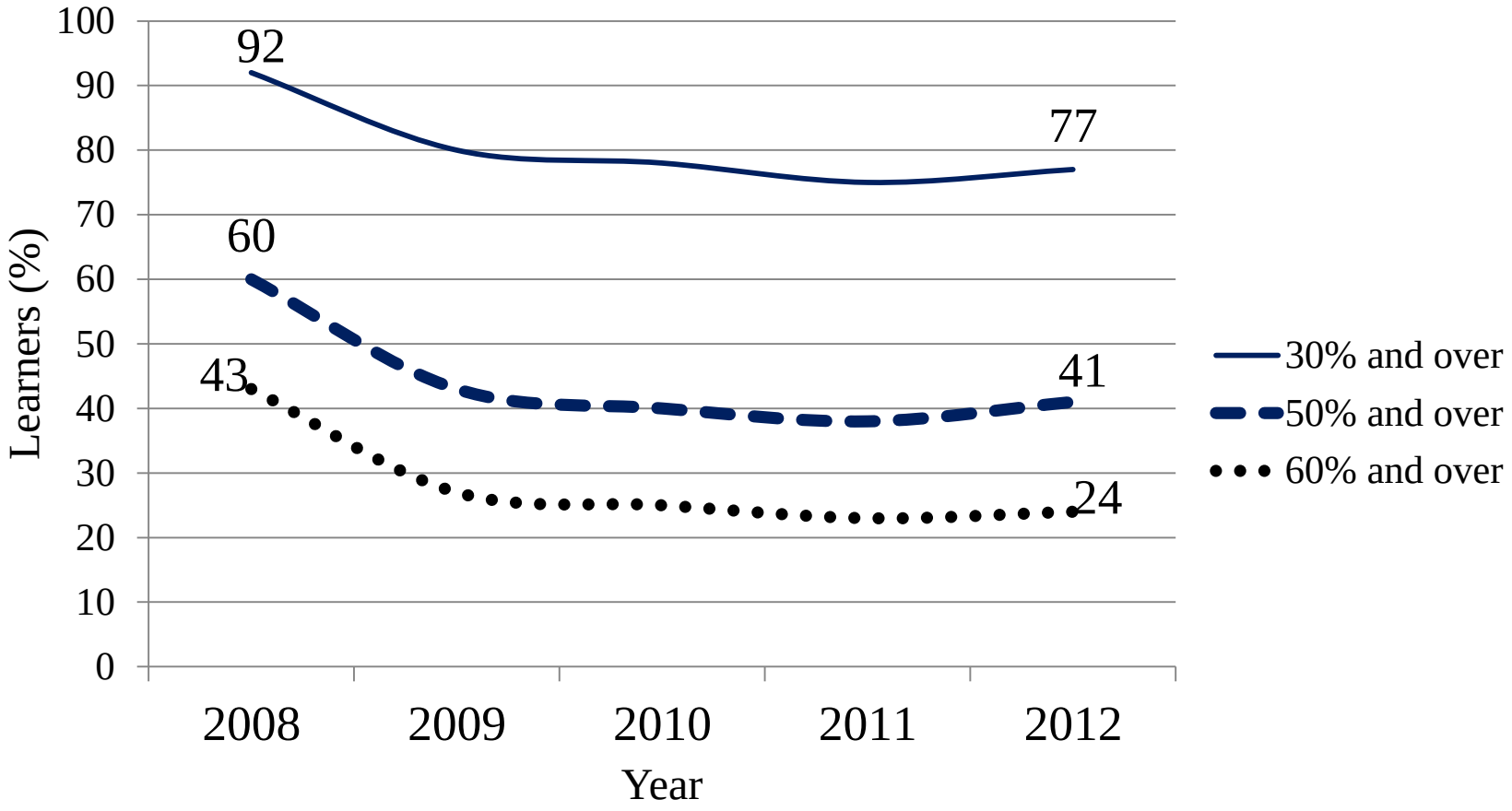
Improving models do not yet account for all of the variance in the dependent variable (learner performance).

Nor do they necessarily account for **changes** in performance over time.

Crouch & Mabogoane (1998): *When the residuals matter more than the coefficients.*

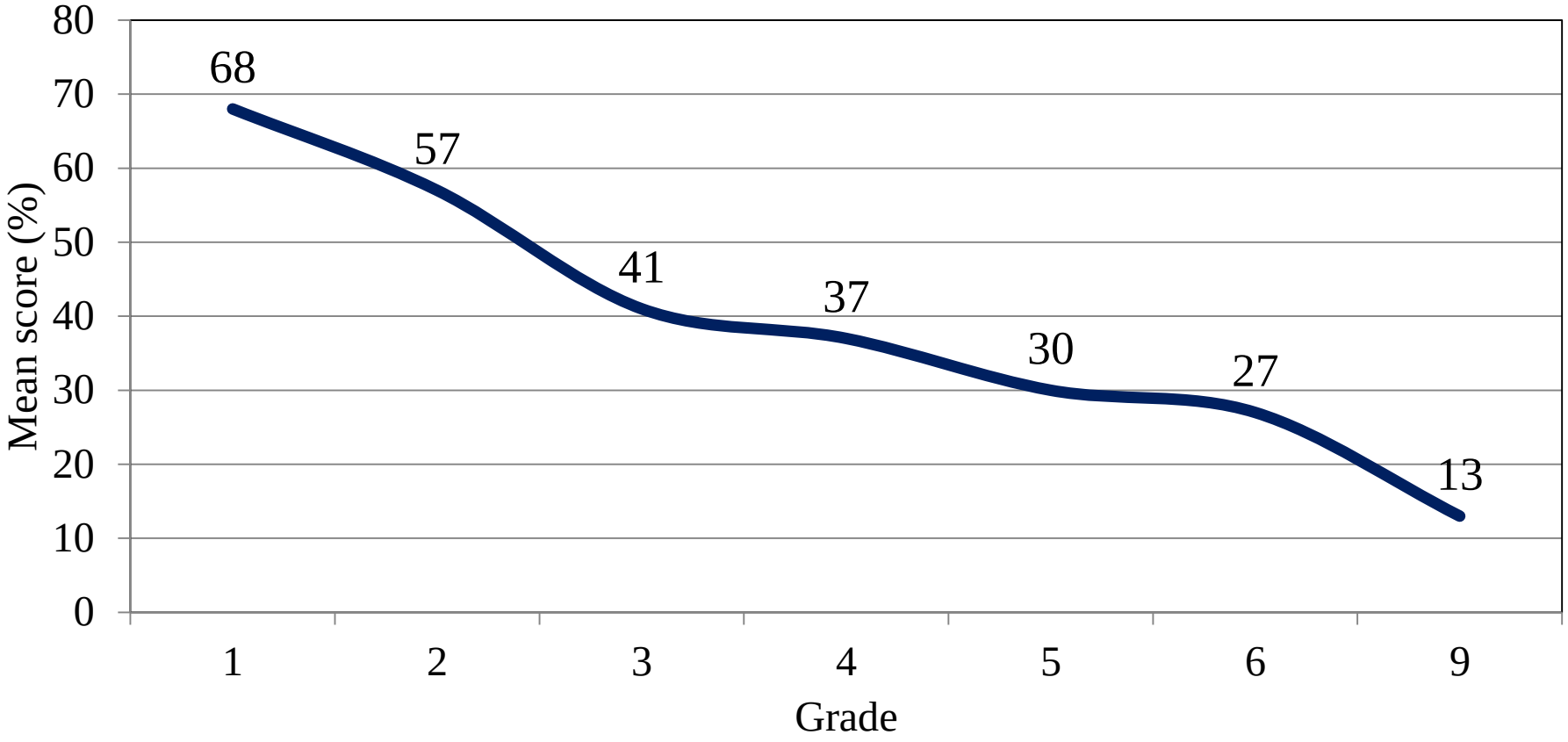
There are still residuals - what are the missing independent variables?

NSC: Learner scores at different levels of quality in 72 high-performing public schools: 2008-2012: MCP



Source: Author using data analyzed by Mouton (2014)

National mean mathematics score: ANA: 2012



The current models clearly explain significant **differences** between high- and low-performing schools.

BUT they cannot explain why performance is **declining** in schools in which none of the independent variables are evident.

There are variables (extrinsic, teacher and instructional quality) that have not been adequately operationalized.

Currently, most useful quantifiable indicator of Quality is Opportunity to Learn. Reeves (2005). Hoadley (2012).

Deaton (2010): a RCT without an understanding of underlying mechanisms – and without the successful handling of heterogeneity

and exogeneity - does not necessarily have a higher epimistic value than other methods.

Randomization of treatment assignment ensures that independent variables are exogenous.

Typical problem for quasi-experimental designs. Danger of a False Positive = low predictive power beyond the local context for larger and more diverse populations. Cautious about generalizing for Policy.

Heterogeneity of treatment is typically problematic.

If treatments differ in scale and/or type in project group, to what model of theoretical, or even operational, mechanisms do impacts on sample means, standard deviations and statistical regressions refer?

Typical problem for consortium-based interventions. Danger of a False Positive or Negative with regard to the impact of the theoretical model. If there is one – may be a number of models.

Deaton welcomes recent trends in development experimentation away from the evaluation of “projects” and toward the evaluation of “theoretical mechanisms”.

Development research is confronted by the problem of simultaneous causality in which the effects of inputs are disguised by effects running in the opposite direction.

Example: Extent of effect of a national curriculum on learner performance is dependent in the first place on the degree of curriculum coverage within the system.

Coverage is *extrinsic* to the *actual inherent value* of *any* model of mechanisms but it very clearly affects performance of learners exposed to test instruments which *assume* that they have covered the whole curriculum.

Again, to what model do sample means, standard deviations and statistical regressions refer?

Danger of False Negative.

Evidence-Based Research and Development

There is a difference between theory and theology. The characteristic feature of a provisionally accepted theory is **Predictive Power**.

Merton (1949) “Middle Range Theory” limits the meaning of theory to hypothesized theoretical mechanisms from which empirical predictions can be derived that are capable of falsifiability.

There are no grand explanatory paradigms with powerful predictive capability in sociology from which to derive and guide *continuing experimentation toward an increasingly accepted, cumulative and comprehensive* theory of social behaviour.

Growing recognition that there are a variety of potentially useful research methods, besides a RCT, that are not experimental, econometric, or even necessarily quantitative, that can be used in combination, is a significant advance in developmental research.

Research methods can range from the wholly quantitative and experimental to the non-experimental and qualitative, even ethnographic, dimensions of the question at hand.

The nature of the information that we are interested in at any stage of an extended programme of developmental research

And the intended use of this information

Determines the methods through which it can be most effectively and usefully collected, used and interpreted.

Phases of an evidence-based programme of research in educational development.

Phase I Exploratory non-experimental study, including reviews of the literature and existing data, resulting in (a) the statement of a causal hypothesis and (b) the development of an operational iteration of the causal hypothesis.

Phase II Randomized Controlled Trial of Model with second post-test: Internal Validity and Maintenance of Effect

Phase III Non-experimental longitudinal replication at large scale in routine system. External Validity.

Phase IV Benchmarking. Social Significance. External Validity.

The Primary Mathematics Research Project: 2004-2012

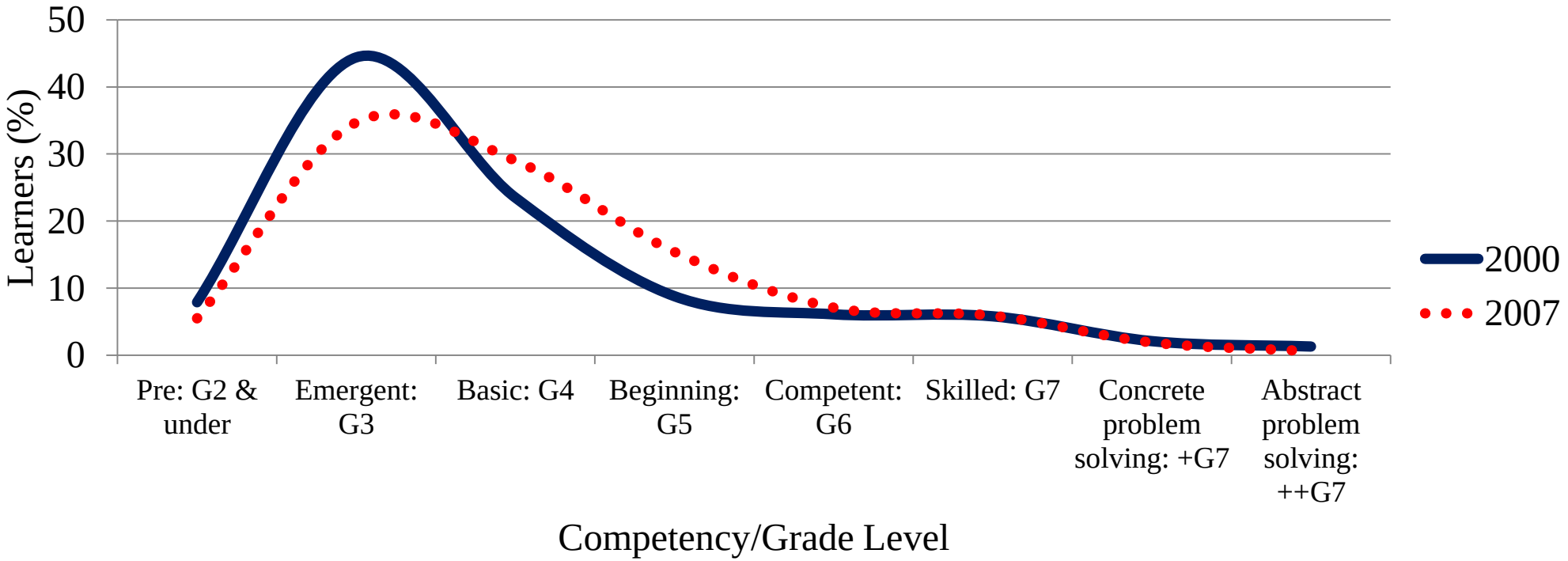
PMRP starts from a known problem that requires *understanding*, *explanation* and *remediation*.

PHASE I: EXPLORATORY STUDY

Strategic Objective

To develop an operationalized Model of Mechanisms, based on theory (deductive) and on the available evidence of different types (inductive), that is capable of sustaining the causative analysis of the impact of the theoretical Model on learner performance.

Grade 6: Distributions of learners by SACMEQ competency levels and equivalent South African grade standards: 2000 and 2007



Is some improvement, from 52.3% to 40.2% at an **emergent** level or below - **BUT** over 80% are below minimum expected standards at both points.

Proportion of learners who did not achieve minimum expected standards for the grade in which they were enrolled (%)

NSE	SACMEQ		TIMSS
Grade 6: 2005	Grade 6: 2000	Grade 6: 2007	Grade 8: 2003
81	85	85	82

- over a seven year period
- three rigorous studies
- using four different instruments
- administered to four different national samples
- at two different grade levels

Showed that around **80%** of learners in Intermediate Phase and beyond are below their minimum expected competency level.

Grade levels refer to the Age Cohort – not the Competency Cohort - of learners.

All schools, even the highest-performing state schools, have become, to one extent or another, multi-grade schools.

Teachers are faced with learner mathematical competencies that can range from the innumerate to the genuinely proficient in the same class.

Equally, learners at a Grade 3 level of competency may be expected to benefit from content instruction and materials pitched two, three or even four, grade levels higher than their *actual* competence level.

Implications for Model Design

Cognitive, conceptual and operational deficits are cumulative and become insurmountable, especially in subjects that are vertically demarcated like mathematics and science. [Spaull (2103) and Simkins (2013)]

Gaps between **assumed** and **actual** grade competence levels have become far too large to be bridged by cosmetic gestures like ‘mixed-ability groups’ or, more recently, ‘differentiation’.

Vygotsky: Zone of Proximal Development. [Chaiklin (2003)]

A mathematical learning programme should be based on a cyclic diagnostic system (feedback) that controls the exposure of learners to

content at, or just above, their existing level of cognitive and conceptual competence and proceeds from there.

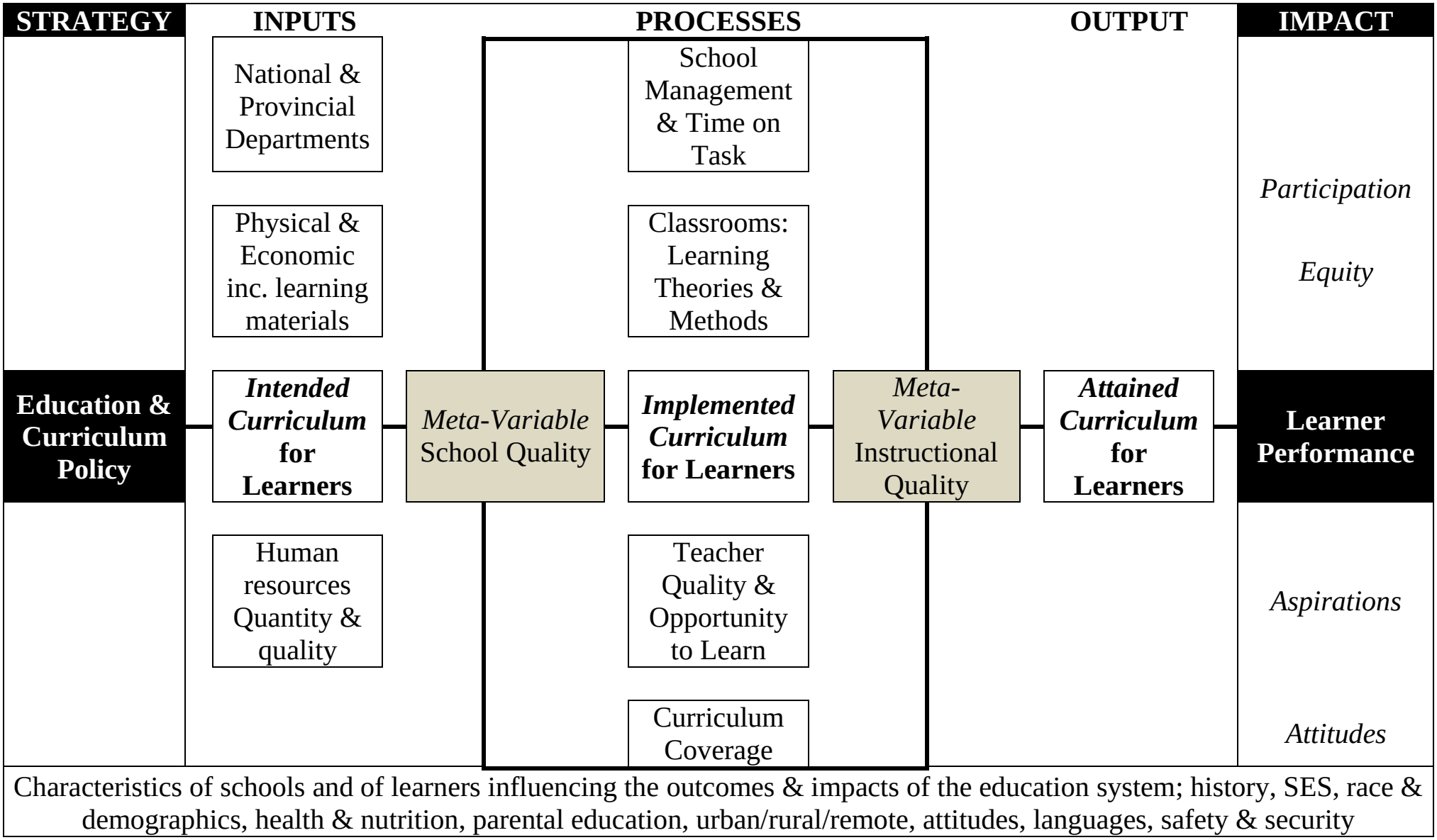
Not just a problem in low-performing schools

Survey of 100 teachers in 64 schools with highest NSC performance: most significant barriers to improving the quality and the quantity of matriculation passes. (MCP)

Prior education of learners: existing backlog in mathematical content

'Severity' ranking of 73% from the whole sample with 62 teachers rating the problem a 'very significant' or 'significant'.

Literature Review: Why is Learner Performance so Poor?



Analysis of existing data derived from:

- six different quasi-experimental impact evaluations
- using the same mathematics test instrument
- conducted in a total of 154 schools
- in 35 Districts in all nine Provinces
- with 7,028 randomly selected learners
- between 1998 and 2004.

Quantitative pre, mid and post-test scores + biographical data

Qualitative data from 1,398 teacher interviews and 1,329 lesson observations.

Raw **primary** data from 4,256 completed test instruments including rough workings as learners actually attempted to solve problems.

Summary of trends in mathematics scores: change in mean score between pre and post-testing (%)

Study Area	Period	G5	G7
National	2000-2003	+1.7	+1.9
Gauteng, KZN, E. & W. Cape	2002-2004	-1.2	-2.9
Gauteng, KZN & W. Cape	1998-2000	-4.2	-3.1
Eastern Cape	1998-2000	-2.1	-0.1
Mpumalanga	1999-2001	-1.1	-2.3

No evidence of any significant improvement of scores.

Analysis of rough workings was very illuminating.

Around 80% of Grade Five and 60% of Grade Seven learners used the simple counting of single unit markings (‘sticks’) to solve problems to one degree or another. 38% and 11%, respectively, relied *exclusively* upon this (lack of) method.

Poor performance is caused by a national inability to perform calculations.

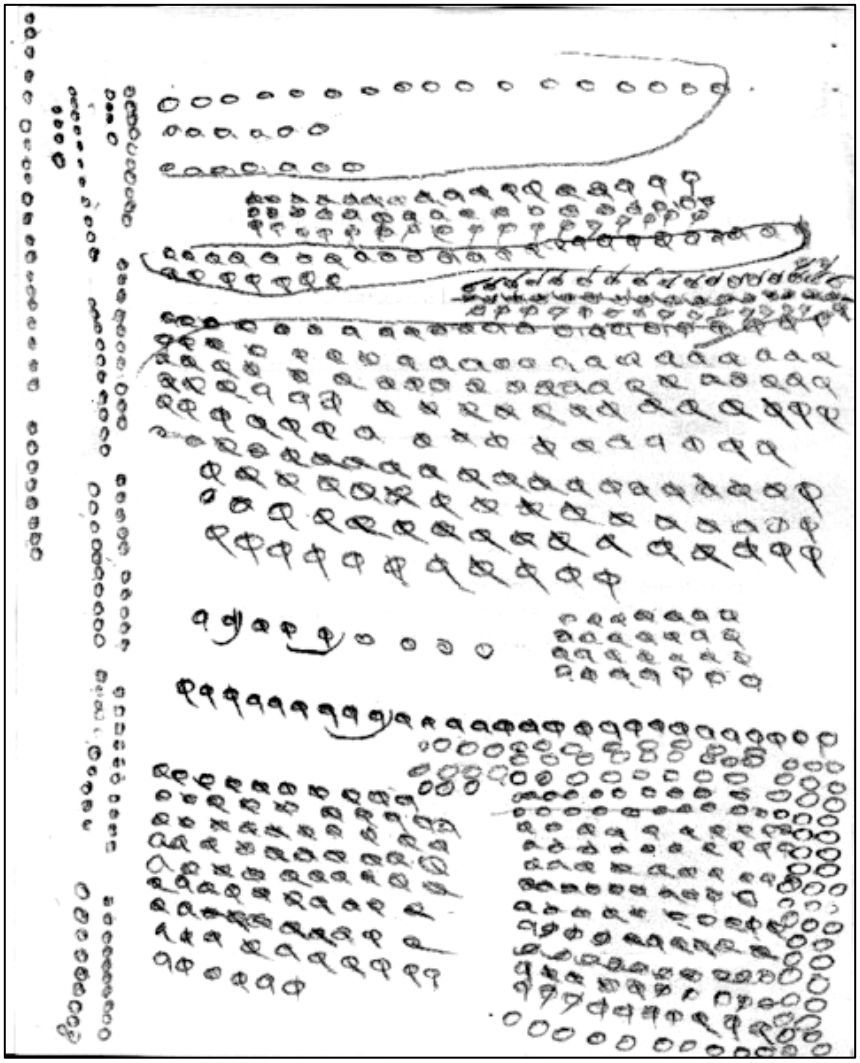
The majority of problems are attempted through the use of simplistic pre-mathematical counting methods in a ‘base-one’ number system in which every number is understood only as the addition or subtraction of another unit to the previous.

Learners do not understand the base-ten number system or the concept of place value and, consequently, lack the procedural capability to use ‘borrowing’ and ‘carrying’ in their attempts to solve problems.

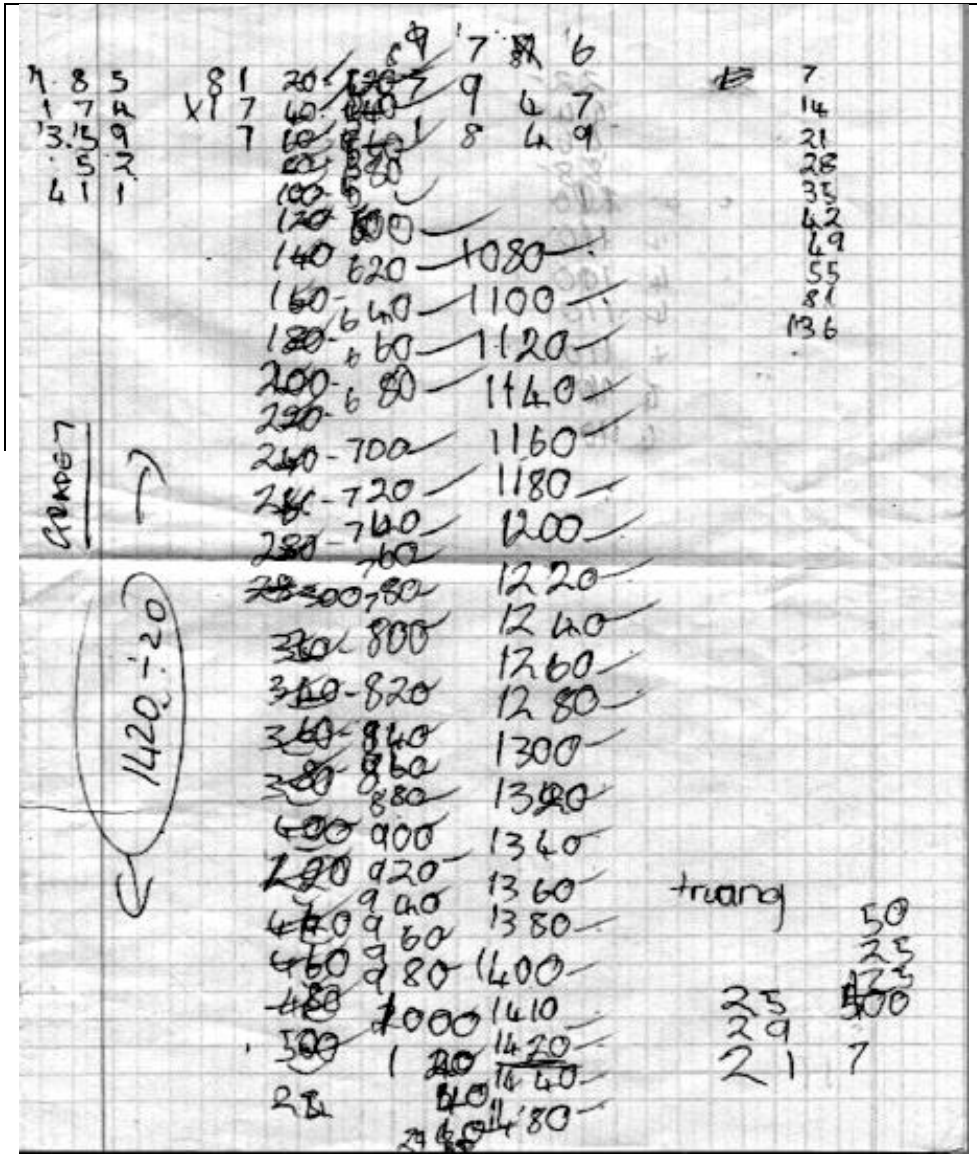
Learners are unable to rapidly and accurately retrieve from memory any learned information about numbers.

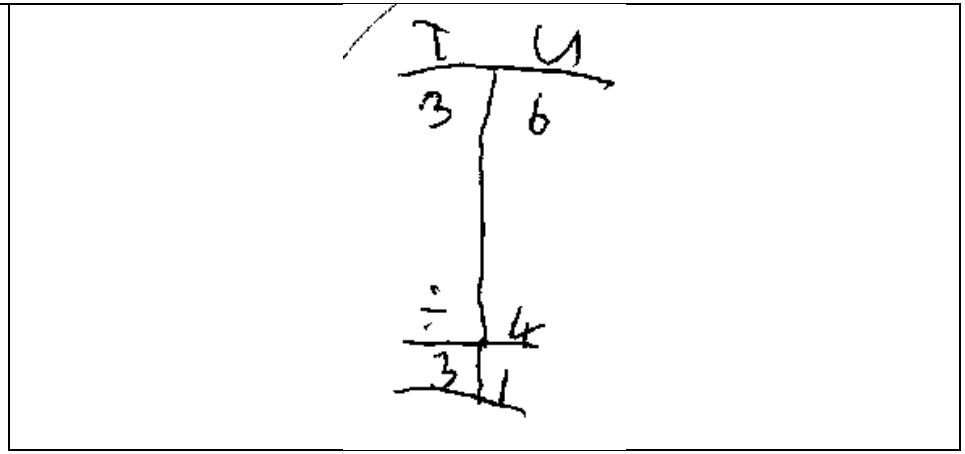
How do children solve problems?

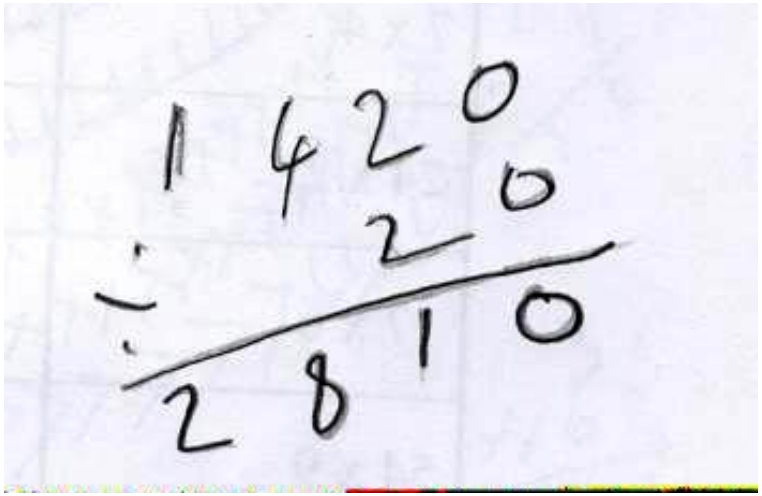
Grade Five Script: Pure Unit Counting



Grade 7 scripts: skip counting and attempts to calculate







How are children taught mathematics?
Teacher Interviews and Lesson Observations

Phrases like:
Learners learn on their own, learners teach each other, educators must facilitate not teach, no rote learning, no spoon feeding, no learner fails, learners learn at their own pace, no repetitive busy work, learners must be active, facilitate their learning, learners must discover knowledge for themselves in groups
Assumed near-universal currency in both INSET programmes and interviews with teachers and departmental officials across the country between 1998 and 2004.

Lesson observations provided consistent qualitative evidence that the problem of learner underperformance was caused by the application of ineffective teaching and learning practices in classrooms.

The discouragement of sustained and direct instruction of content by teachers in favour of a much more limited role as a facilitator of ‘discovery’ or ‘active’ learner-centred learning in collaborative groups.

The virtual disappearance of memorization, consistent drill and regular extensive practice of newly learned content.

Learners routinely worked in a distracting environment comprised of multiple factors when dealing with unlearned content knowledge

Complex tasks that were not understood, methodological practices/roles (group work) they had to adopt, multiple tools (e.g. ruler & protractor) or sources of information that were provided but not taught or explained, noise and movement by disengaged learners.

Whatever curriculum planners may have intended, the lesson observations and teacher interviews provided consistent evidence of how their theoretical and methodical assumptions had been interpreted in practice in classrooms around the country since 1998.

Why did teachers do what they did? OBE Training

CHARACTERISTICS OF PROGRESSIVE AND TRADITIONAL TEACHERS

PROGRESSIVE	TRADITIONAL
1. PUPILS ENCOURAGED TO BE AUTONOMOUS.	1. PUPILS ARE RESTRICTED IN FREEDOM / INDEPENDENCE.
2. ACTIVELY PARTICIPATE IN LEARNING/TEACHING SITUATION.	2. KNOWLEDGE TRANSMITTED PASSIVELY.
3. CHILD CENTRED LEARNING.	3. TEACHER CENTRED LEARNING
4. ASSERTIVE.	4. SUBMISSIVE.
5. TEACHERS KNOW PUPILS BY THEIR FIRST NAMES.	5. TEACHERS KNOW PUPILS BY THEIR SURNAMES OR OTHER NAMES.
6. PUPILS LEARN TO THEIR FULL POTENTIAL.	6. LEARNING IS LIMITED TO THE SYLLABUS / CURRICULUM.
7. KNOWLEDGE SORTED OUT THROUGH DISCOVERY LEARNING.	7. TEACHERS POURS OUT KNOWLEDGE AND PUPILS 'RECEIVE'.
8. TEACHER ASKS PROBING QUESTIONS TO DEVELOP UNDERSTANDING.	8. ENCCOURAGE ROTE LEARNING AND MEMORISING.
9. TEACHER IS AWARE OF PUPILS' STRENGTH AND LIMITATIONS.	9. TEACHER CAPITALISES ON PUPILS' LIMITATIONS AND RIDICULES CHILDREN.
10. ENCOURAGES LEARNING FOR LIFE.	10. ENCOURAGES LEARNING CERTIFICATION.
11. OPENS UP TO WHAT PUPILS BRING AS WELL AS ASKING QUESTIONS.	11. TEACHER HAS A SET AGENDA AND IGNORES PUPILS INTERESTS.

Conclusion of the Exploratory Study

The use of methodological forms based on an understanding of 'learner-centred' approaches that imply a theory of teaching and learning based on the 'active discovery' or 'construction' of 'new' knowledge by the learner with the teacher as a 'facilitating supporter' - rather than as a direct instructor of established content and practice - is a very significant cause of poor learner performance in mathematics.

The use of these forms, as typically practised in the majority of South African schools, actively militate against the conceptual and cognitive quality of the **Opportunities to Learn** mathematics that are provided to learners.

Age-based grade cohorts mean that the exposure of learners to content at, or near, their actual competence levels (zone of proximal development) must be handled internally through a diagnostic sub-system.

Statement of the Research Hypothesis

A *literature review* of international educational and neurocognitive research following the Exploratory Study. Kirschner, et al (2006); Abadzi (2006); Steel and Funnel (2001); Sweller (1989), Nykiel-Herbert (2004)

There is a causative relationship between the use of methodological forms used by teachers in classrooms and levels of learner performance.

Evidence for the greater effectiveness of direct and guided instruction is explained in the context of our current knowledge of neurocognitive architecture and of cognitive load.

‘Discovery’, ‘active’ and ‘collaborative’ methodological forms overwhelm the short-term processing capacity of learners and inhibit the efficient transfer of information to long-term memory (learning).

Learning deficits in a cognitively and conceptually hierarchical subject such as mathematics are cumulative and result ultimately in dismal performance at matriculation level and beyond.

Therefore

The Model of Mechanisms is based on a cognitive theory of learning requiring the direct and guided instruction of content by teachers.

Attention is paid to the characteristics of working memory and the importance of the relationship between short-term processing capacity and long-term memory.

The instructional programme is operated through a diagnostic system to control exposure of learners to appropriate cognitive and conceptual content.

Application of this Model will result in:

- Significant improvements in learner performance.
- Significant improvements in problem solving methods *from counting to calculating*.

Basic Design Principles of the Operational Treatment Programme

Information provided in small and understandable ‘amounts’ that are logically connected to the previous and following pieces of information. These pieces of information/skills are practiced or repeated immediately after instruction while short term memory still retains the ‘template’ required for doing so.

Only once sufficient of these pieces of information, skills and templates are assimilated into long term memory is it possible to develop progressively and cumulatively the conceptual schema (***implicit memory***) that are the basis of mathematical fluency, accuracy and creativity, and of all higher-order mathematics itself.

The Model was operationalized in the form of a programme of instruction consisting of teacher and learner materials for a period of **14 weeks** with **70 lessons** for **Grades 3, 4, 5 and 6** based on the Assessment Standards of the NCS for the then-Learning Outcome One: Numbers Operations and Relationships.

By the end of Foundation Phase, learners should have mastered basic computational methods and have discarded simplistic pre-mathematical counting.

Learner Workbooks provide multiple daily exercises dealing with content based on the AS for all levels from Grades 3 to 6. Teacher Manuals provide scripted lessons and are keyed to the different weeks and days of the exercises in the learner workbook.

Diagnostic testing for each operational procedure distinguished each learner by actual grade-competency level.

A teacher cannot present content at four different levels of learner competence during each lesson.

He/she instead presents direct instruction in and explicated a specific skill, competence or operational procedure.

Learners subsequently referred to the section of the workbooks that allows them to practice this skill or procedure at their own level of competence.

PHASE II: A RANDOMIZED CONTROLLED TRIAL

INTERNAL VALIDITY: DOES THE MODEL 'WORK' OR CAN IT BE REFUTED?

Operated in 2007 in three circuits of the Vhembe District in NE Limpopo.

Almost all of the schools are in remote-rural areas. Representative of the South African schools which achieve the lowest impacts on performance.

40 schools were invited by the Limpopo Department of Education. Schools were randomly ascribed to project (treatment) and control (counterfactual) groups; 20 schools in each.

Each project school received sets of diagnostic tests, teacher manuals and learner workbooks for two complete classes (grades 4 &6). Each Teacher attended a 1.5 day workshop in the use of the operational treatment programme.

Each teacher received 2 monitoring & support visits – one just after initiation and one prior to Week 7 of the 14 week programme.

Objective was to understand and increase degree of homogeneity of treatment received by learners at classroom level.

Two impact measures

- Whole sample irrespective of degree of programme coverage
- Post-trial sub-group analysis: project schools which presented 11 weeks (80%) of the programme.

Post-trial sub-group analysis violates a basic principle of a RCT and random selection. Produces a different project group that may not be equivalent to the original, or to the control, at baseline.

BUT significant heterogeneity of treatment **also** violates a basic principle of a RCT!

Discarding scores of low-coverage project schools actually *improved* pre-trial equivalence by reducing the difference of the pre-test mean scores by 1%.

Although belonging to the whole project group at baseline did have a statistically significant effect on the dependent variable, the R^2 of 0.023 indicated that the effect was not strong explaining **2.3%** of the variance in the pre-score. Once low-coverage schools were discarded, the effect was *weaker* with a R^2 of 0.0154 explaining **1.54%** of this variance.

Impact: Whole project group irrespective of coverage

	Mean gain (% points)	Mean growth of pre score (%)
Project	+13.6	+77.3
Control	+3.1	+21.2
Impact	**+10.5**	**+56.1**

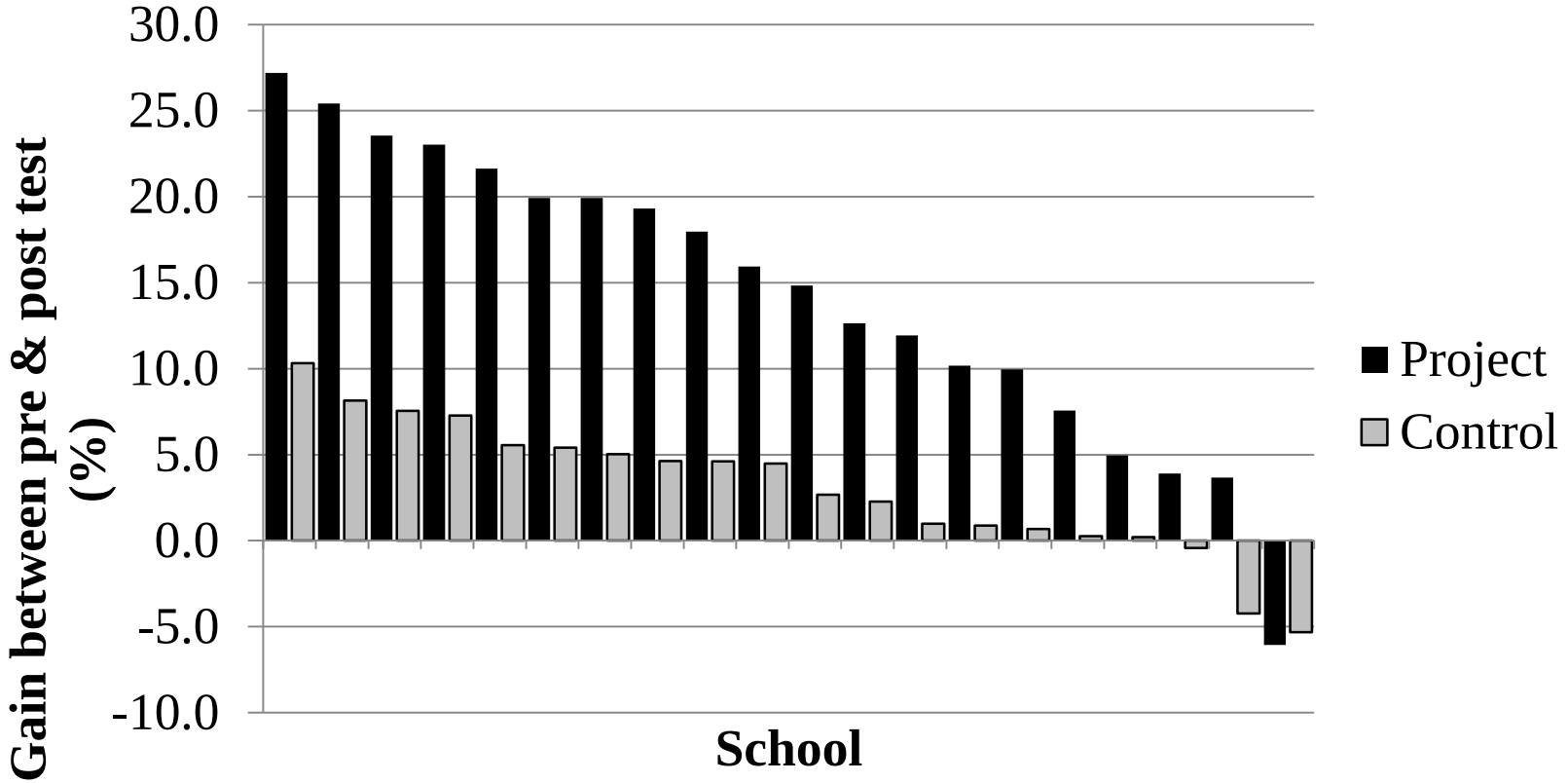
$F(1, 3030) = 600.2$, $p<0.05$ with a t of 24.5, a coefficient of 10.54, and a R^2 accounting for **16.5%** of the variance in gains.

Impact: High-coverage project group

	Mean gain (% points)	Mean growth of pre score (%)
Project	+18.6	+111.9
Control	+3.1	+21.2
Impact	**+15.5**	**+90.7**

$F(1, 2334) = 1{,}295.4$, $p<0.05$ with a $t = 36.00$, a coefficient of 15.87 an a R^2 accounting for **35.7%** of the variance in gains.

Comparison of gain scores of schools according to rank position in project and control groups irrespective of coverage by project group



14 of the 20 project group schools obtained a higher gain than any of the 20 control schools.

Project Group: Gain in score between pre and post testing by pre-score learner quintiles

Quintile	Mean gain in score	Std. Dev.	Freq.
1	16.2	11.5	161
2	17.0	11.7	239
3	18.3	12.8	134
4	19.7	13.2	169
5	22.4	14.5	159

Significant gain was obtained by all pre-score quintiles. Range 6.2%. The highest gains were obtained by the highest quintiles, as were the highest standard deviations. Range 3.

Significant impact was spread across the whole sample – schools and learners.

The hypothesized causal relationship between the theoretical Model of Mechanisms and significant improvements in learner performance has not been refuted. We accept, provisionally, that the Model ‘works’.

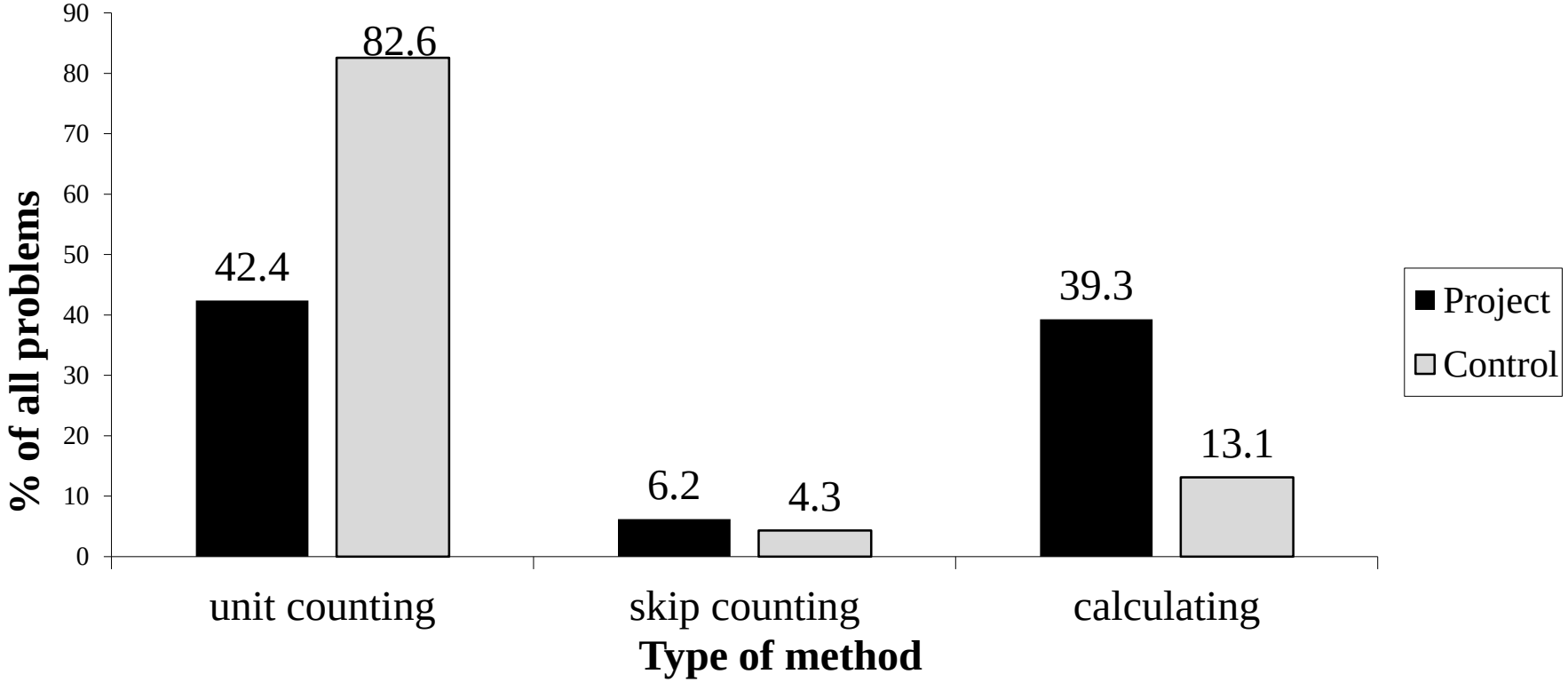
The RCT provides a range of predictive impact values for future iterations of the theoretical Model under similar conditions.

Between:
+10.5% and 15.5% in percentage points
+56.7% and 90.9% in the growth of the pre-score

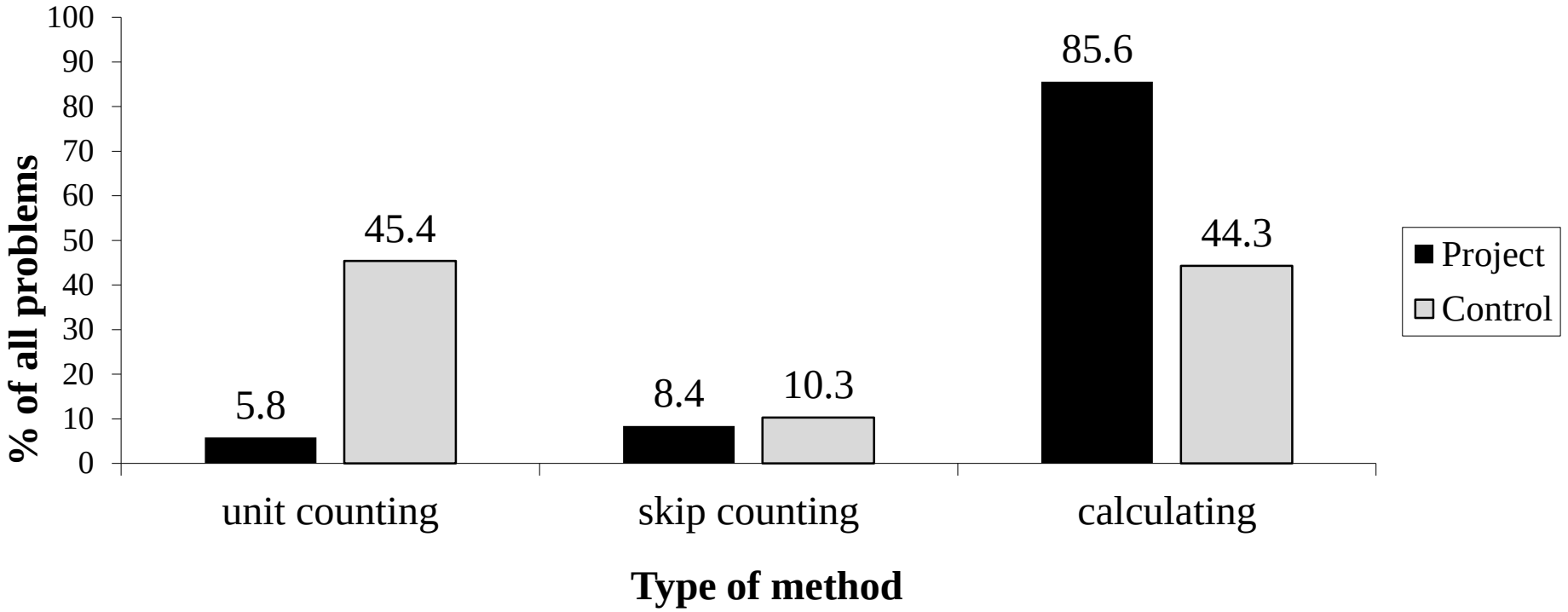
The lower values apply to a population with an unknown degree of treatment coverage. The higher values apply to a population that has covered at least 80% of the treatment programme.

Impact Indicator II: Problem-solving Methods used by Learners

Frequency of methods used by percentage of all problems: Grade 4

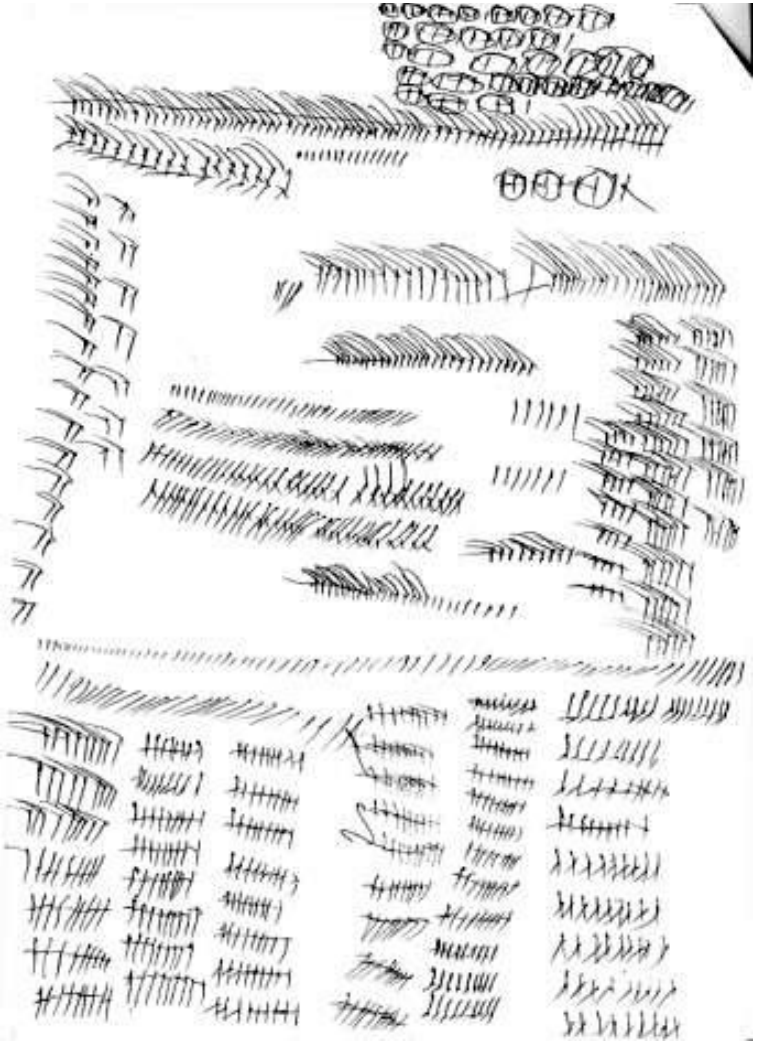


Frequency of methods used by percentage of all problems: Grade 6

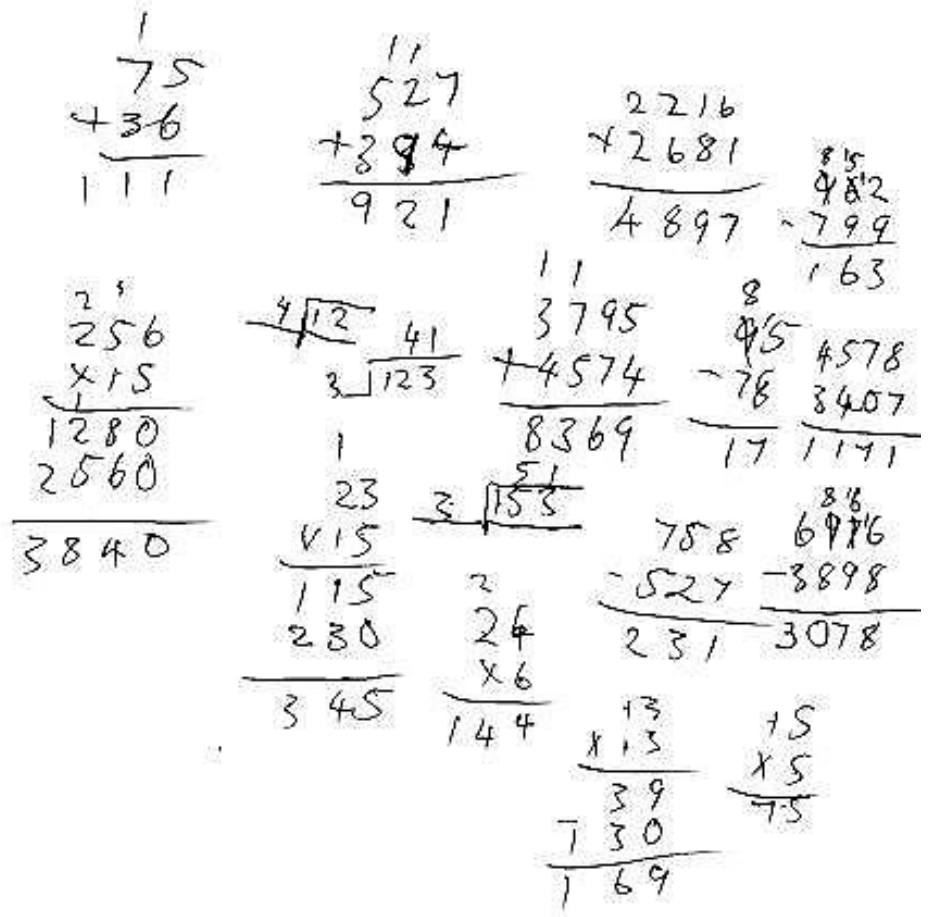


As the Model predicted, improvements in learner performance between pre and post testing are matched by large differences in problem solving methods by post testing.

Grade 6 Scripts

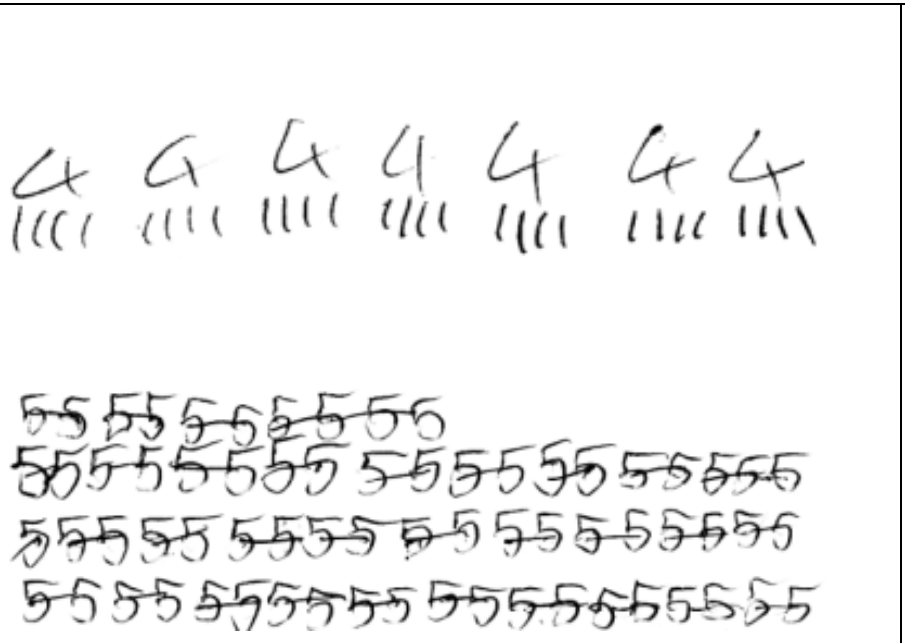
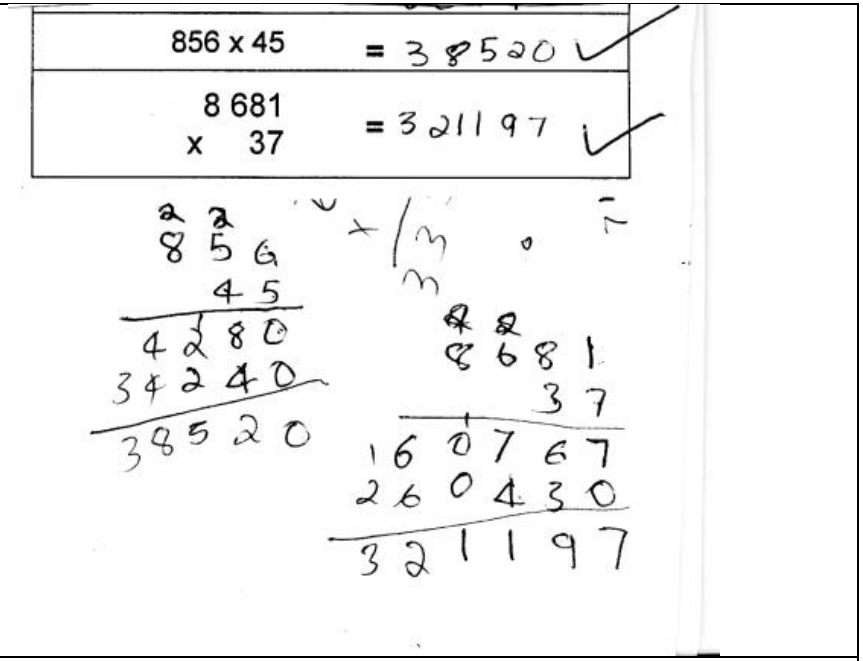


Control: Counting



Project: Calculating

Grade 6 Scripts

	
Control: 7 x 4 and 67 x 5	Project: 856 x 45 and 8,681 x 37

Extension to the RCT

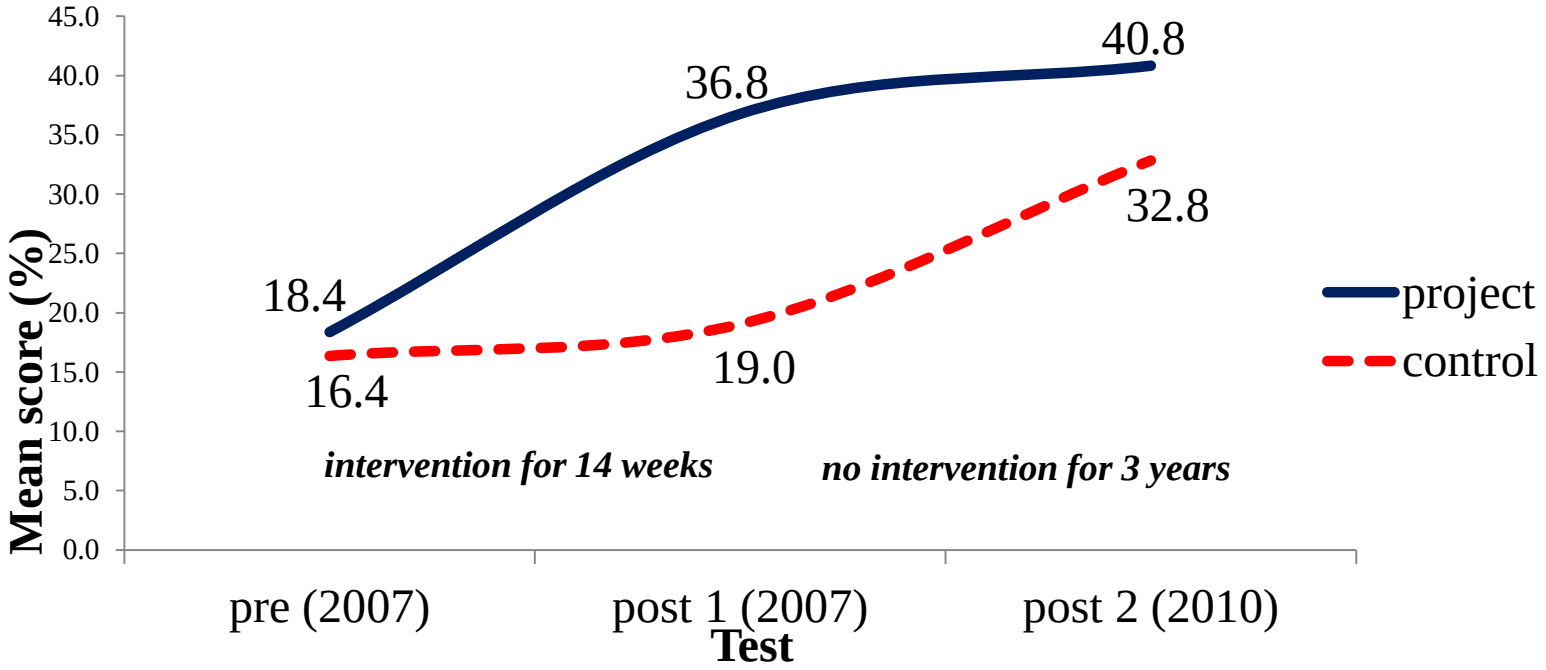
Maintenance of Effect
2007 to 2010

After 14 weeks of intervention treatment in 2007, schools in the project group returned to the routine schooling received by those in the control group.

In 2010, a second post-test was administered to the same project/control cohort that was tested at baseline in Grade 4 in 2007.

These learners were in Grade 7 in 2010 and were tested on the RCT Grade 6 instrument.

Mean scores of RCT cohort project and control groups between pre-test (2007) and post-test 2 (2010)



3 years after intervention-end, the control group had not reached the same score as the project group after, at least, 11 weeks of exposure in 2007.

Learners in the project group learned more in 14 weeks than those in the control group in 3 years.

PHASE III:

Replication at a large scale within the routine system under different conditions.

External validity of the Model: Non-experimental study

"RCTs of 'what works', even when done without error or contamination, are unlikely to be helpful for policy … unless they tell us something about why the program worked … scientists who run … experiments are likely to do so more carefully and conscientiously than would the bureaucrats in charge of a full scale operation.

In consequence, there is no guarantee that the policy tested by the RCT will have the same effects … (after large-scale replication) … as in the trial even on the subjects included in the trial or in the population from which the trialists were selected.

For an RCT to produce 'useful knowledge' beyond its local context, it must illustrate some general tendency, some effect that is the result of mechanism, that is likely to apply more broadly … it is the combination of mechanism and context that generates outcomes and … without understanding that combination, scientific progress is unlikely." (Deaton 2010: 448-449)

In January 2010, the programme was extended to all of the schools in the 5 circuits of the Malamulele Cluster in Vhembe.

Class sets of diagnostic tests, learner workbooks and teacher manuals were provided to all of the participants for Grades 4, 5 and 6. After a one-day session for schools on using the programme, the researchers withdrew from management and support of the programme.

- 125 schools
- 370 teachers
- Over 20 000 learners
- 5 Circuit Managers
- 125 Principals/SMT members
- 3 Vhembe District Mathematics Advisors
- 5 Circuit PMRP Teacher Committees
- 1 Area PMRP Teacher/SMT Committee

All schools selected, Two whole classes at each school selected.

Two main operational objectives of the 2010 to 2012 study were to:

- ❖ Establish the impacts obtained by the intervention treatment on learner performance. *Would it obtain the same, or similar, levels of impact on learner performance as the 2007 RCT on a much larger scale and without external support?*

- ❖ Establish and understand the requirements for *transferring* the routine operation, management and support of the intervention treatment *to the department* without significant loss of programme quality and, hence, lower levels of impact on learner performance.

The RCT obtained a predicted effect for a replication of the intervention in a similar environment/context of between +10.5% and +15.5% over one cycle of treatment. (In percentage points)

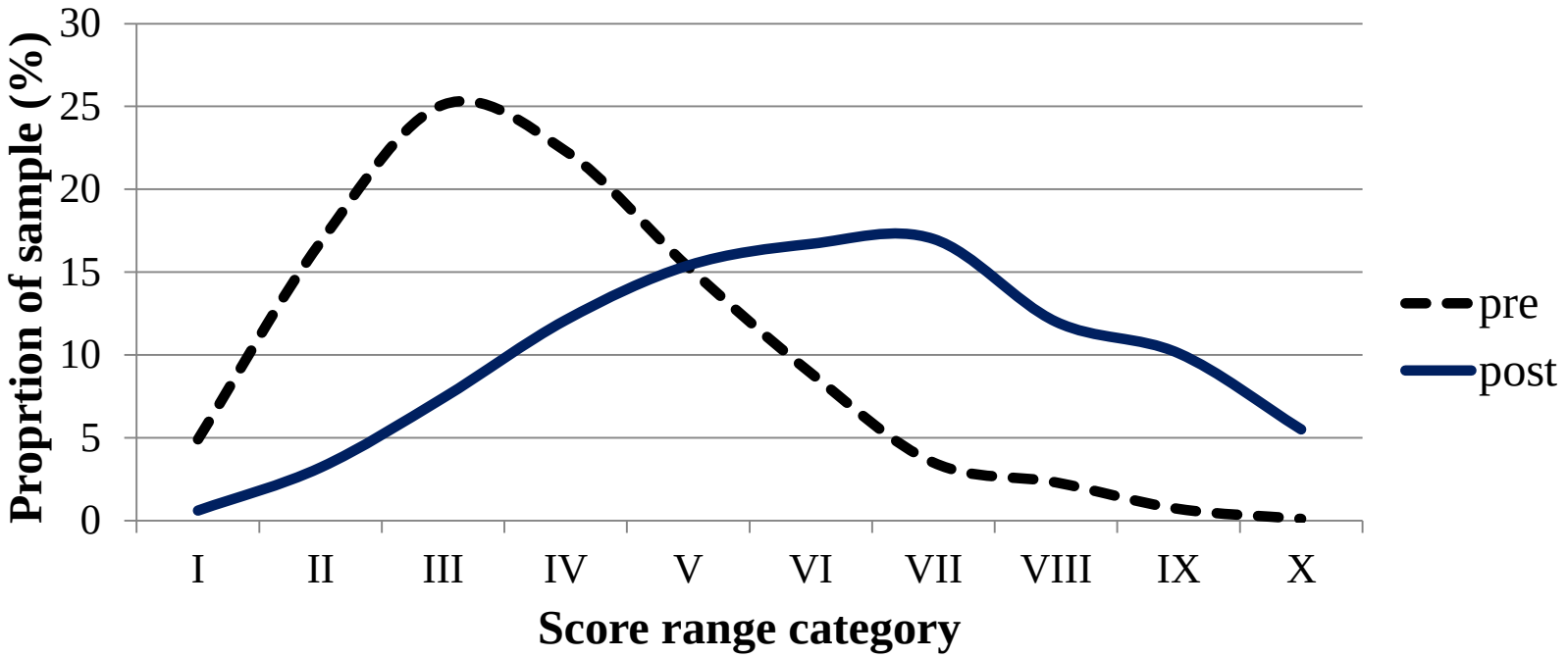
The impact estimation for the replication study obtained over a similar period of one cycle between pre and mid testing in 2011 was **+12.4%**, almost at the mid-point of the predicted range.

By post testing in 2012, the cumulative difference between the pre and post scores had doubled to **+24.8%**.

These cumulative gains were spread consistently across each of the five districts within a range of 5.3%: a high of +26.5% and a low of +21.2%.

As an increase of the baseline score, the RCT predicted range was between +56.1% and 90.6%. The baseline score increased by **+77.3%**, again almost at the mid-point of the predicted range

Distribution of cohort scores between pre and post testing



Positive improvements in performance were obtained across the whole sample in all of the score range categories.

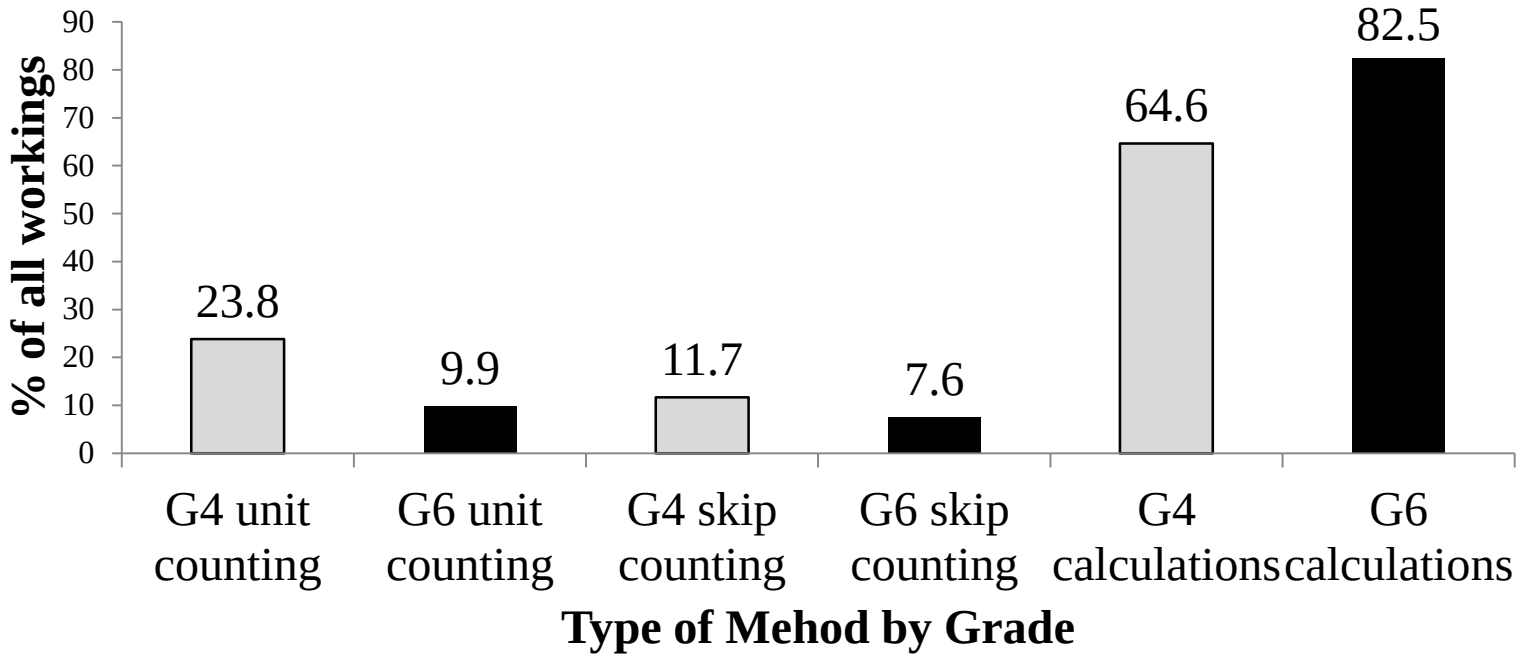
Gain in learner score by post testing by pre-test quintile

Pre-test quintile	Mean gain	Std. Dev.	Frequency
1	+23.1	14.4	292
2	+26.4	15.0	261
3	+27.3	14.7	242
4	+25.9	14.3	230
5	+21.6	14.3	238

While gain is concentrated in the three middle quintiles to some extent, the range across all five is very low at 5.7%.

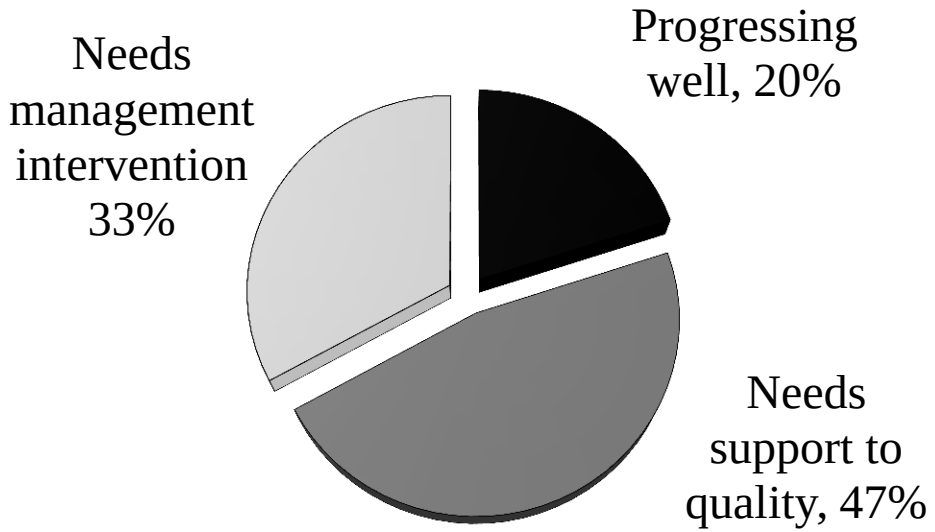
Learners of all pre-existing ability levels achieved significant gains in performance as a result of exposure to the programme intervention.

Frequency of problem-solving methods used at post-testing by percentage of all workings



A significant increase in learner performance scores has been accompanied by a significant increase in the ability of learners to transcend pre-mathematical counting methods in favour of the use of whole numbers in conventional calculations.

Status of programme delivery by quality category in all circuits: 2010 (% of schools)



Consistent evidence that the mechanisms used by the model had begun to spread in the replication study area before initiation of the replication study. Partly as a result of the 'leakage' of the

methodological forms of the PMRP programme itself, and partly as an effect of other departmental initiatives, especially the National Foundations for Learning Campaign and the Common Work Programmes of the District.

These programmes, to one degree or the other, used similar mechanisms as the PMRP, and their mutually-reinforcing conceptual integration was very strongly influenced by the appointment of a Circuit Manager who had participated in the RCT as the District GET Mathematics Coordinator in the period between the two studies.

He was also appointed as the District Departmental Project Manager for the PMRP between 2010 and 2012.

His role and effect was critical to the use of the Model itself.

Most significantly from a theory building point of view, the replication study found that there was no uniformity or consistency in the operational application of the PMRP programme itself as a sequential and paced course of instruction, or even in the degree to which the whole programme of instruction was completed.

The PMRP programme was, instead, almost universally used as a generic support to the achievement of the objectives of all of the departmental programmes including CAPS.

The one critical distinguishing factor which applied to all 125 schools was the application of the PMRP diagnostic test to direct learners initially to the appropriate level of cognitive and conceptual capacity.

The evidence of the replication study is that the effectiveness of the mechanisms operationalized in the PMRP intervention programme is not, in fact, dependent on a specific programme (project) iteration.

Nonetheless, as the RCT subgroup analysis shows, the effectiveness of the model in achieving a higher degree of improvement is strongly influenced by the degree to which an ***organized course of instruction based on these mechanisms is presented and completed***.

PHASE IV: Social Significance & Equity

Speculative and Indicative: Directions for future research?

Comparison of proportion of PMRP Grade 6 learners at Grade 5 competence level or below with the comparable figures obtained by the NSE and SAQMEC (% of sample)

NSE National	NSE Limpopo	SACMEQ National	PMRP Pre	PMRP Post
2005	2005	2007	2010	2012
80%	90%	80%	96%	61%

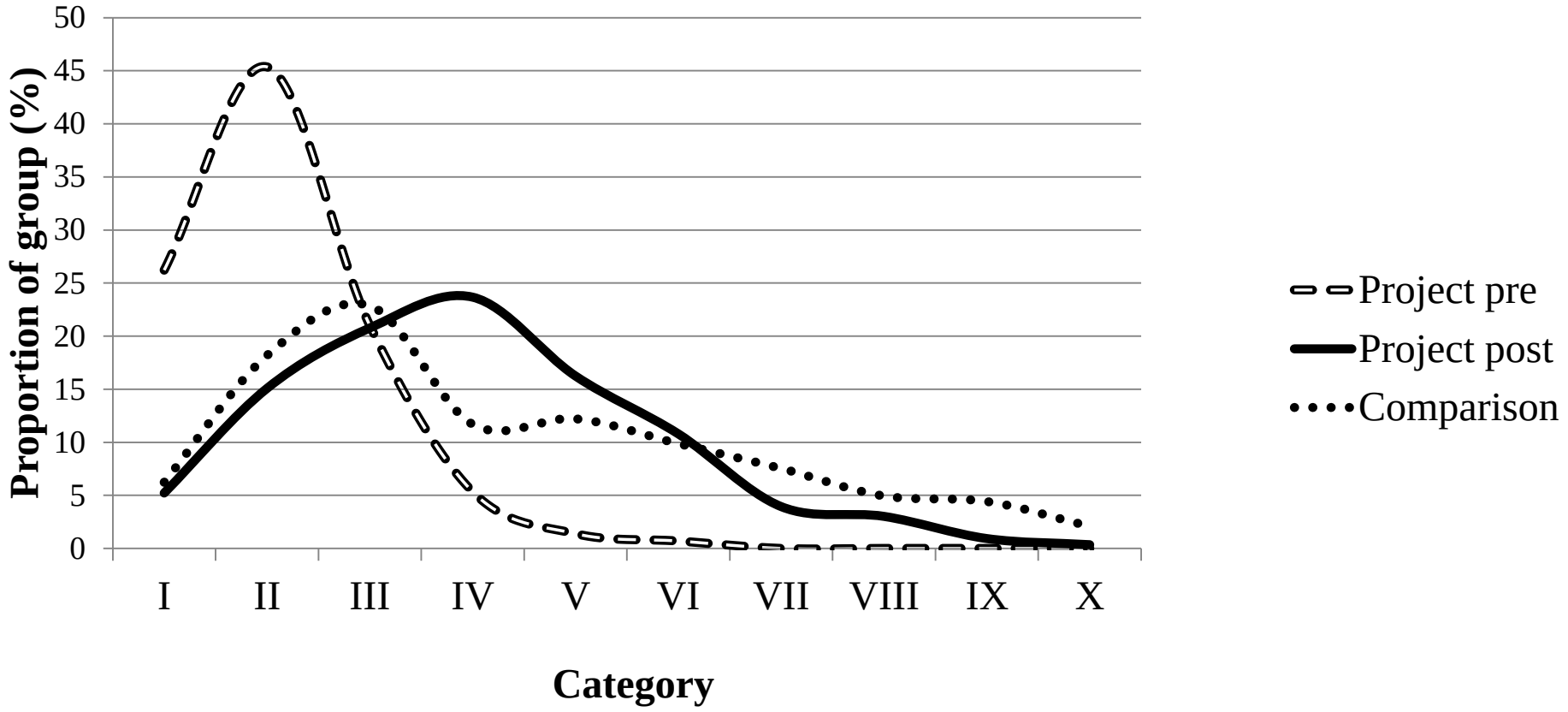
The PMRP schools were all in rural and remote-rural schools. These schools have the lowest performance levels in the country. Leaving aside the W. Cape for demographic reasons, Gauteng state schools obtain the highest performance levels.

To provide a plausible proxy comparator between these two extremes, all learners in Grades 4 and 6 in two former-Model C schools in Johannesburg were tested in 2010 using the PMRP test instruments.

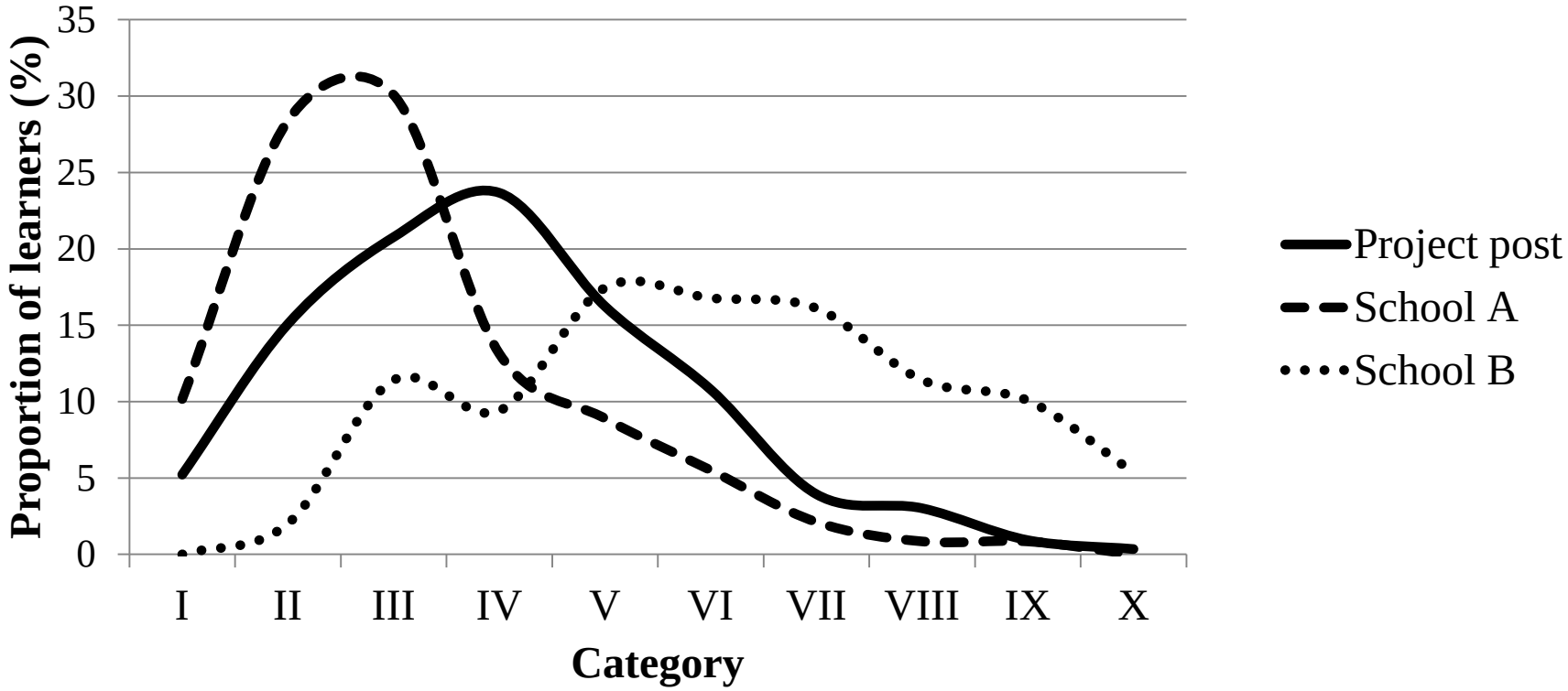
One is a long established suburban middle class school with historically excellent results. Around half of all learners are black, coloured or Indian. Virtually all learners live in the feeder suburbs with parents who are employed and have been enrolled from Grade 1.

The second comparator is a long established working class school in an industrial area with historically good results. Virtually all learners are black. Just over half live in informal settlements, a sizable minority is dependent on social grants and around 20% have arrived from rural areas in the last few years.

Benchmarking: Comparison of distribution of learner scores by range categories of the RCT project group (2010) and the aggregated comparison sample



Comparison of distribution of scores by range categories in each comparison school and RCT project group (2010)



The evidence supports the conclusion that the statistically significant impacts estimated by the PMRP are also potentially socially and educationally significant.

The PMRP project group performance was lower at baseline than either of the comparator schools. By post-test, it had surpassed school B and was closing the gap with school A.

Most significant implication is that it may not take as many years as feared before we see both increasing levels of ***quality*** *outcomes in mathematics and an increasing level of* ***equalization*** *across the system.*

Different performances of the comparator schools = a great deal of outcome inequality even in well-resourced and managed schools.

CONCLUSION

There is consistent evidence of different types that an instructional programme in mathematics that makes use of the Model of Mechanisms is likely to achieve intended impacts on learner performance at large scales.

The degree to which this impact is achieved depends on:

- The degree to which an organized and coherent operational instructional programme is provided to learners. Maximize Opportunity to Learn.

- The degree to which departmental and school management and monitoring succeeds in covering the complete instructional programme. Maximize Opportunity to Learn.