

**The Prevalence of Cheating in National
Standardised Assessments in South African Schools**

*Applying the Jacob and Levitt (2003) method to the Annual
National Assessments of 2013*

By Lauren Ross



Internal Supervisor: Professor Ingrid Woolard

(University of Cape Town)

External Supervisor: Professor Nicholas Spaull

(University of Stellenbosch)

University of Cape Town

Department of Economics

Thesis presented in partial fulfilment of the requirements for the degree of Master of Commerce
in the Faculty of Economic and Management Sciences at the University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

ABSTRACT

The Prevalence of Cheating in National Standardised Assessments in South African Schools

Applying the Jacob and Levitt (2003) method to the Annual National Assessments of 2013

By Lauren Ross

This paper builds on the underlying framework of Jacob and Levitt (2003) and further work by Gustafsson (2014) to determine the prevalence of cheating in South African schools using a nationally representative sample (Verification ANA) of the Annual National Assessments of 2013. This research forms part of a broader research project on 'binding constraints in education' promoted by the Programme to Support Pro-poor Policy Development (PSPPD), housed at Stellenbosch University. The adapted methodology is verified in order to ensure that the suspicious string indicators as discovered by Jacob and Levitt are indeed indicative of suspicious behaviour when applied to the South African ANA. At a national level, the data suggests that cheating or suspicious behaviour is likely to be prevalent in up to 10 percent of schools with respect to Mathematics and Language in Grades 3 and 6. The manner in which schools behave suspiciously varies significantly by province, subject and measure. As many as 37% of primary schools in the Eastern Cape, 26% of primary schools in KwaZulu-Natal and 24% of primary schools in Limpopo show some evidence of cheating in Grade 3 mathematics, compared to just 0% to 2% of primary schools in the Western Cape and Gauteng. Similar extremes are noted in the Grade 6 results. These results suggest that the mere act of assessment and measurement induces behavioural distortions such as gaming behaviour even in the absence of high-stakes.

2016



Plagiarism Declaration

1. I certify that I have read and understand the Commerce Faculty Ethics in Research Policy. <http://www.commerce.uct.ac.za/Downloads/Faculty%20Ethics%20in%20Research%20Policy.PDF>
2. I certify that I have read the General Rules and Policies Handbook (Handbook 3) regarding Student Rules of Academic Conduct: RCS1.1 to RCS3.2 and Rules Relating to Examinations G20.1 to G22.2.
3. I certify that I have read and understand the document, “Avoiding Plagiarism: A Guide for Students.”
4. This work has not been previously submitted in whole, or in part, for the award of any degree in this or any other university. It is my own work. Each significant contribution to, and quotation in, this dissertation from the work, or works of other people has been attributed, and has been cited and referenced.
5. I authorise the University of Cape Town to reproduce for the purpose of research either the whole or any portion of contents in any manner whatsoever.

Student number:	RSSLAU005
Student name:	Lauren Lee Ross
Signature of student:	<div style="border: 1px solid black; padding: 5px; display: inline-block;">Signed by candidate</div>
Date:	14 July 2016

Acknowledgements

This paper uses the theory of Jacob and Levitt's (2003) paper, "Rotten Apples: an investigation of the prevalence and predictors of teacher cheating," and later adaptive work by Gustafsson (2014) to provide an estimate of the prevalence of cheating in the South African Annual National Assessments of 2013. This research forms part of a broader research project on 'binding constraints in education' promoted by the Programme to Support Pro-poor Policy Development (PSPPD) housed at the Research on Socioeconomic Policy (ReSEP) group at Stellenbosch University. It is as part of this research that the Verification ANA 2013 data was obtained and Gustafsson's preliminary paper entitled "A check on item-level data patterns in the 2013 ANA associated with possible cheating" was shared. Unfortunately this thesis was not completed by the time the PSPPD final report was being written and thus it was not included in the final report. I am grateful to the Presidency, the Department of Basic Education, the European Union, Stellenbosch University and the ReSEP group for allowing and enabling analysis of this nature on the V-ANA 2013 data.

Gustafsson's preliminary paper entitled "A check on item-level data patterns in the 2013 ANA associated with possible cheating" adapts the principles of Jacob and Levitt's paper to suit the ANA data and discusses these methods rigorously in regard to the Grade 3 Mathematics assessments. The rigorous verification of the methodology employed by Gustafsson ensures that the cheating indicators as discovered by Jacob and Levitt are indeed indicative of suspicious behaviour when applied to the South African ANA. It is Gustafsson's Stata Ado-File which is applied in this paper to examine the prevalence of cheating, in a broader context including Grades 3 and 6 Mathematics and Language in South African schools. I am grateful to Martin Gustafsson for allowing me to use his do-file.

On a more personal level I am grateful for my supervisor, Nic Spaull, and co-supervisor, Ingrid Woolard, for their administrative help, continued guidance and support. In particular I thank Nic for his enthusiasm on this topic and for pushing me to attack a topic I would otherwise have thought too difficult.

Furthermore to my amazing family and friends, I could never have achieved any of this without your moral support and unconditional love. Dad, thank you for never questioning my academic decisions and always pushing me to reach this goal. You are my biggest inspiration.

Table of Contents

Chapter 1: Introduction	1
Chapter 2: Literature Review	3
2.1 The Importance of Education.....	3
2.2 Differing Degrees of Accountability	3
2.3 Assumptions of New Accountability Systems	7
2.4 Gaming the System.....	8
2.4.1 What is Gaming Behaviour?.....	8
2.4.2 Strategic Forms of Gaming Behaviour	8
2.4.3 Unintentional Strategic Behavior.....	10
2.5 Intentional Cheating and Rotten Apples	12
2.6 The Case for South Africa	13
Chapter 3: The Data	16
3.1 Introduction to the ANA Data	16
3.2 Completeness of the Data.....	17
3.3 Values on Class within Grade	18
Chapter 4: Methodology	19
4.1 The Construction of Cheating Indicators.....	19
4.1.1 Indicator 1 (Unexpected Test Score Fluctuations)	19
4.1.2 Indicator 2 (Suspicious Answer Strings).....	20
4.2 Identification Strategy and Assumptions.....	24
4.3 Cheating Measures and Overall Levels of Performance	26
Chapter 5: Data Analysis and Findings	28
5.1 Summary Statistics: V-ANA Academic Performance.....	28
5.2 Detailed Analysis of Grade 3.....	29
5.2.1 Summary Statistics for ANA Cheating Measures.....	29
5.2.2 Analysis of Cheating Measures by Province	31
5.2.3 Patterns within Individual Schools	35
5.2.4 Summary.....	46
5.3 Detailed Analysis of Grade 6.....	46
5.3.1 Summary Statistics for V-ANA and SACMEQ Cheating Measures	47
5.3.2 Analysis of Cheating Measures by Province	51

5.3.3	Patterns within Individual Schools	59
5.3.4	Summary.....	73
5.4	Grade 9	73
5.5	Other Interpretations.....	74
Chapter 6: Conclusion and Recommendations for Further Work.....		75
Reference List.....		77
Appendix.....		82

List of Boxes

Box 1: Grade 3 Mathematics Items on which Suspicious Schools Outperformed the Average ..	40
Box 2: Grade 3 Mathematics Items over which Suspicious Schools Underperformed	40
Box 3: Grade 3 Language Items on which Suspicious Schools Outperformed the Average	45
Box 4: Grade 3 Language Items over which Suspicious Schools Underperformed	45
Box 5: Grade 6 Mathematics Items on which Suspicious Schools Outperformed the Average ..	65
Box 6: Grade 6 Mathematics Items over which Suspicious Schools Underperformed	66
Box 7: Grade 6 Language Items on which Suspicious Schools Outperformed the Average	72
Box 8: Grade 6 Language Items over which Suspicious Schools Underperformed	72

List of Figures

Figure 1: An Example of V-ANA Cheating Measures Percentile Plots	26
Figure 2: Illustration of Strange School EC 301	36
Figure 3: Illustration of Strange School EC 302	37
Figure 4: Illustration of Strange School KZN 303	38
Figure 5: Illustration of Non-Strange School KZN 304.....	39
Figure 6: Illustration of Strange School KZN 305	41
Figure 7: Illustration of Strange School NC 306.....	42
Figure 8: Illustration of Strange School KZN 307	43
Figure 9: Illustration of Non-Strange School MP 308	44
Figure 10: Illustration of Strange School MP 601 (A).....	61
Figure 11: Illustration of Strange School FS 602 (S)	61
Figure 12: Illustration of Strange School EC 603 (A)	62
Figure 13: Illustration of Strange School EC 604 (S)	63
Figure 14: Illustration of Non-Strange School WC 605 (A).....	64
Figure 15: Illustration of Non-Strange School KZN 606 (S).....	64
Figure 16: Illustration of Strange School FS 607 (A).....	68
Figure 17: Illustration of Strange School EC 608 (S)	68
Figure 18: Illustration of Strange School KZN 609 (A)	69
Figure 19: Illustration of Strange School EC 610 (S)	70
Figure 20: Illustration of Non-Strange School FS 611 (A).....	71
Figure 21: Illustration of Non-Strange School MP 612 (S)	71

List of Tables

Table 1: Details on Items per Test	17
Table 2: Correlation Matrices of Cheating Measures.....	25
Table 3: Summary Statistics on the Academic Performance of V-ANA Schools.....	28
Table 4: Summary Statistics on Cheating Measures for Grade 3.....	29
Table 5: Combinations of High Value Cheating Measures for Grade 3.....	30
Table 6: 90th Percentile Cheating Measure Values by Province for Grade 3.....	31
Table 7: Number of Noted Schools by Province for Grade 3.....	33
Table 8: Percent of Noted Schools by Province for Grade 3	33
Table 9: Number of Schools at least 2 Standard Deviations above the National 90th Percentile for Grade 3.....	34
Table 10: Percent of Schools at least 2 Standard Deviations above the National 90th Percentile for Grade 3.....	34
Table 11: Summary Statistics on Cheating Measures for Grade 6 Mathematics	47
Table 12: Summary Statistics on Cheating Measures for Grade 6 Language.....	48
Table 13: Combinations of High Value Cheating Measures for Grade 6 Mathematics	50
Table 14: Combinations of High Value Cheating Measures for Grade 6 Language.....	50
Table 15: 90th Percentile Cheating Measure Values by Province for Grade 6 Mathematics	51
Table 16: Number of Noted Schools by Province for Grade 6 Mathematics	53
Table 17: Percent of Noted Schools by Province for Grade 6 Mathematics.....	53
Table 18: Number of Schools at least 2 Standard Deviations above the National 90th Percentile for Grade 6 Mathematics	54
Table 19: Percent of Schools at least 2 Standard Deviations above the National 90th Percentile for Grade 6 Mathematics	54
Table 20: 90th Percentile Cheating Measure Values by Province for Grade 6 Language.....	55
Table 21: Number of Noted Schools by Province for Grade 6 Language.....	56
Table 22: Percent of Noted Schools by Province for Grade 6 Language	57
Table 23: Number of Schools at least 2 Standard Deviations above the National 90th Percentile for Grade 6 Language.....	57
Table 24: Percent of Schools at least 2 Standard Deviations above the National 90th Percentile for Grade 6 Language.....	58

Chapter 1: Introduction

Effective educational systems have long been considered an important determinant of economic growth and as such sit at the heart of most development strategies (Lucas, 1988 and Mankiw *et al.*, 1990). Improved educational outcomes reflect an increase in human capital which should result in a positive effect on individual earnings and overall lead to a strong positive effect on economic growth (Barro & Lee, 2010). The focus on improving educational outcomes used to be centered around the expansion of educational attainment but has recently shifted towards theories of improving educational quality as “schooling that does not improve cognitive skills, measured by comparable international tests of mathematics, science and reading, has a limited impact on aggregate economic outcomes and on economic development” (Hanushek & Woessmann, 2008, p.608). In an attempt to provide quality education, many governments have implemented accountability programmes with varied stakes (incentives and sanctions) based on performance measures which consistently assess and monitor student, and sometimes teacher, cognitive skills. A variety of researchers have found that improved performance on these tests is shown to be closely related to individual productivity and earnings (Hanushek, 2002; Lazear, 2003; Mulligan, 1999; Murnane *et al.*, 2000). The substantial magnitude of the impact of cognitive skills on economic returns indicates that “educational quality concerns are very real for developing countries [such as South Africa] and that this aspect of schools simply cannot be ignored” (Hanushek & Woessmann, 2008, p.621).

According to the standards and accountability movement a series of three activities are needed in order to improve education: “first, defining what students should learn (setting standards); second, testing to see what students have learned (measuring achievement); third, making the results count (holding educators and students accountable)” (Loveless *et al.*, 2005, p.7). Currently the South African education curriculum outlines the standards of content using the Curriculum Assessment Policy Statement (CAPS). The implementation and testing of these standards is handled by schools and these schools are held accountable to the Annual National Assessments (ANA) as stipulated by the Department of Basic Education (DBE). It would seem that South Africa has started to put in place the practises necessary to raise the standard of education delivered. There is, however, a growing concern that some strategic behaviour, or cheating, occurs during the implementation of these assessments, allowing schools who are underperforming to go unnoticed and unsanctioned.

The review report of the first ANA undertaken in 2011 titled “Report of the Annual National Assessments of 2011” included some analysis of patterns in the item-level data for Grade 3 that suggested that some level of cheating may occur (Department of Education, 2011). Specifically it was found that, “in three provinces (Eastern Cape, KwaZulu-Natal and Limpopo), the degree of similarity between the responses to the same question by learners in the same school suggests that test administration procedures may not have been properly carried out” (Department of Basic Education, 2011, p.28). The report further states that in some schools these statistical anomalies may be driven by incorrect teaching within individual schools or classes (Department of Basic Education, 2011). Both cheating and incorrect teaching are particularly dangerous in education and hence these topics need to be further explored.

This paper examines the topic using widely tested methods introduced by Jacob and Levitt in the United States (2003). This analysis focuses on the detection of ‘strange patterns’ in the data which points to cheating. These irregularities in the testing process could result from (a) assistance offered by teachers during the test, (b) student copying or (c) adjustments made to test scripts after the assessment. This paper will therefore use the underlying framework of Jacob and Levitt (2003) and further work by Gustafsson (2014) to determine the prevalence of cheating in South African schools using Verification ANA 2013 data. The results emanating from the Verification ANA data are likely to underestimate the prevalence of cheating for the very reason that the schools selected for the verification sub-sample had additional external personnel (from Deloitte) present at the schools (or rotating between the V-ANA schools) in order to monitor the implementation of the assessments. The additional monitoring under these conditions is likely to deter cheating.

Chapter 2: Literature Review

2.1 The Importance of Education

Education is a process which enables human beings and societies to reach their fullest potential. One of the main objectives of education is to increase the human capital of individuals with basic skills such as reading, writing and mathematics. The ripple effect of education, however, is much broader improving the lives of individuals and stimulating the economy through a number of mechanisms. Education increases the efficiency of individuals raising the earning potential of the labour force, fosters democracy creating better conditions for good governance (Barro, 1998 and Bobba & Coviello, 2007), improves health and reduces fertility (Cutler *et al.*, 2006 and D'Addio *et al.*, 2005) and promotes equality (Aghion *et al.*, 1999). Investment in education is therefore a key determinant of a country's development process and as such is often the largest single line item of government expenditure (as is the case in South Africa). As such there is a strong incentive for government to ensure that these resources achieve the outcomes for which they were procured resulting in educational policy debates and continued restructuring of education systems in an effort to improve the accessibility and quality of education. Recent studies, however, conclude that "education investments [only] contribute to faster GDP growth if schooling is effective in raising student learning" (Bruns *et al.*, 2011, p.142). This has led to a shift in the global focus on improving education towards a more stringent accountability system through government policy and international monitoring (Chiang, 2009; Cullen & Reback, 2006; Jacob & Levitt, 2003). Increased accountability should motivate all stakeholders through incentives and sanctions to raise the quality of education provided thereby offering students the means to lead a productive life consistent with sustainable development and effective public participation as set out in Agenda 21 of the United Nations (United Nations Decade of Education for Sustainable Development, 2002).

2.2 Differing Degrees of Accountability

In an attempt to improve the quality of education, many governments have implemented an accountability programme based on performance measures. The shift toward increased accountability aims to provide "adequate checks and incentives to ensure that teachers and school administrators maximise student learning" (Cullen & Reback, 2006, p1). The focus on student performance as a measure of the health of a school diverges from previous thinking

which stressed school budgets, resources and management to ensure provision of a standardised and acceptable level of education (Fuhrman & Elmore, 2004). This shift in thinking comes as a result of mounting evidence which reveals that “what students actually learn, not how many years of schooling they complete, is what counts for economic growth” (Bruns *et al*, 2011, p.4). This point highlights the quality of education offered over the quantity of education attained. There is a substantial body of evidence reflecting the varying degrees of accountability that have been employed by both developed and developing countries, and their relative impacts.

Incentives and Sanctions – The United States of America

The United States of America (USA) employs both incentives and sanctions in motivating stakeholders to improve educational quality. These policies have been widely implemented through the No Child Left Behind Act of the Bush administration in 2001 (Chiang, 2009). Under this Act schools administer a series of standardised math, literacy and language assessments to every student in grades three to eight. These tests are then used to grade schools A through F depending on student performance. The grading of schools is made public knowledge and schools repeatedly failing to meet performance standards are penalised accordingly which may include the redirection of funding, possible replacement of school leadership and or staff, state takeover of operations and in severe cases school closure (Figlio, 2006). Under this Act thirteen states use public reporting as the primary mechanism for attaching stakes to performance, nineteen states acknowledge exceptional performance with monetary or non-monetary rewards and nineteen states have specific policies enabling them to reconstitute schools and change staff (Goertz & Duffy, 2001). The public disclosure of school grades in some instances may increase the stakes associated with the accountability programme through the effect this information has on other markets. Figlio and Lucas (2000) provide evidence that housing markets are highly sensitive to school report cards, signalling the attractiveness and potential of an area.

This new educational environment provides significant incentives for underperforming schools to alter the way in which they deliver their services and, in the United States increased accountability has typically resulted in a positive effect on the performance measures of students (Carnoy & Loeb, 2002; Hanushek & Raymond, 2004; Springer, 2008). Furthermore a cross-state analysis of varying degrees of accountability shows that students in high-accountability states average significantly higher scores on the National Association of Educational Progress (NAEP) math test (Carnoy & Loeb, 2002). This is the hope of the accountability programme, that threatened schools would restructure finances and resources in order to improve efficiency and raise performance standards. This is exactly the story outlined by Chiang (2009) who estimates

that the threat of sanctions from being graded as failing induced lasting student improvement of approximately 0.11 standard deviations in reading and math in the following year due to increased school expenditure on instructional and curricular development, reading coaches and specialised consultants, teacher training and technology.

Incentives and Information – Brazil (Pernambuco)

Brazil's accountability initiative involves the dissemination of information and monetary rewards for improved performance. Pernambuco is plagued with very low educational outcomes, ranking last out of Brazil's 26 states on the national education quality index (Ferraz & Bruns, 2011). In 2008 Pernambuco became the third state to implement a group pay-for-performance system that uses large monetary rewards to incentivise school personnel to meet annual school improvement targets, the effects of which were studied by Ferraz and Bruns (2011). Improvement targets are set for each school, for grades 4, 8 and 11 on the SAEPE (Pernambuco State Learning Assessment) which tests Maths and Language annually. Information regarding current school performance and future school targets was sent to school principals and made available to the public via school notice boards (Ferraz & Bruns, 2012). All schools that reach at least half of their set targets are rewarded with a proportional bonus which could average over one month's salary dependent on the number of schools reaching their targets. The significant incentives created by the design of this system resulted in an almost instant improvement in education with a 17 percent increase in the number of schools achieving their targets from 52 percent in the first year to 79 percent in the second year. The rewards received averaged approximately 1.8 months and 1.4 months of a recipient's salary in the first two years of the programme respectively (Ferraz & Bruns, 2012). In addition, schools with more ambitious targets achieved greater improvements in student outcomes and classroom observations carried out just prior to the end-of-year student achievement exams revealed that "schools whose teachers spent more time on instruction were much more likely to achieve a bonus" (Ferraz & Bruns, 2012, p.4).

Information only – India (Andhra Pradesh) and Kenya

In the Indian state of Andhra Pradesh an information only campaign provides experimental evidence on the effectiveness of low-stakes accountability programmes. The majority (80 percent) of children in rural Andhra Pradesh attend one of over 60 000 government-run schools (Pratham, 2008). With a high student to teacher ratio, strong teacher unions and a lack of disciplinary action, teacher attendance and performance is low (Kremer *et al.*, 2005). Against this backdrop a set of 100 schools were randomly selected in the 2005-6 school year to receive baseline and endline academic tests, absolute and relative written diagnostic feedback to teachers

and low-stakes monitoring of classroom practices. Improved information is meant to inform teacher practice and should increase intrinsic motivation by “focussing attention on student learning levels and improving their ability to set and work towards goals” (Muralidharan & Sundararaman, 2010, p.2). Evaluation of this programme found significant impacts on the effect of information and low-stakes monitoring on teacher behaviour. In treatment schools “teachers were more likely to be observed actively engaged in teaching, using textbooks, engaging students, and using the blackboard” (Bruns *et al*, 2011, p.38). In spite of increased effort by teachers, students in treatment schools fared no better than those in control schools on endline assessments (Muralidharan & Sundararaman, 2010). Whilst increased information had no effect on student outcomes, it did induce a positive change in teacher behaviour.

The effects of information provision on student performance vary widely by study. In a number of other contexts such as Liberia, Pakistan and Uganda, information for accountability has led to increased student performance (Bruns *et al*, 2011). However closer to home, a recent randomised educational intervention in Kenya has shown that information provision has no discernible effect on either private or collective action (Lieberman *et al*, 2014). In this study, public citizen activism and the private actions of those in the household are evaluated between those households that did and did not receive the randomised information package. The information disseminated to parents included the performance of their school-aged children as well as suggestions on materials and strategies that may improve their child’s learning. This initiative was shown to be unsuccessful as, “parents that received the informational treatments were no more likely than other parents to take action at school or in the public sphere to improve the quality of their children’s schooling or to adopt behaviours at home that might have a positive impact on their children’s learning” (Lieberman *et al*, 2014, p.70).

International Comparisons and Benchmarks

Many countries do not specifically have an educational accountability programme but may be covered by international studies. There are a number of benchmark assessments such as the Programme for International Student Assessment (PISA), the Progress in International Reading and Literacy Study (PIRLS) and the Trends in International Mathematics and Science Study (TIMSS). Locally, there is the Southern and eastern Africa Consortium for Monitoring Education Quality (SACMEQ). This study covers the conditions of schooling and quality of education in a host of southern and eastern African countries using reading and math assessments (Moloi & Chetty, 2010). SACMEQ, TIMSS and PIRLS have been particularly important for South Africa as, until recently, the matric school-leaving examinations were the

only resource available to evaluate educational progress. These studies monitor and report on the global standard of education and are useful in promoting public accountability through debate and pressure for reform (Bruns *et al*, 2011).

2.3 Assumptions of New Accountability Systems

There are five main assumptions underlying the new accountability systems such as those seen in the United States. According to O'Day (2004) in *Redesigning Accountability Systems in Education* these new systems assume that:

- Performance, or student achievement, is the key value or goal of schooling, and constructing accountability around performance focusses attention on it.
- Performance is accurately and authentically measured by the assessment instruments in use.
- Consequences, or stakes, motivate school personnel and students.
- Improved instruction and higher levels of performance will result.
- Unfortunate unintended consequences are minimal.

In application, however, these assumptions frequently fail. Whilst the first two points can be argued to fail on theoretical or philosophical grounds, the last three, which are of particular interest, can be said to fail in practice due to their human element. Consequences, or stakes, are incorporated into accountability programmes in order to incentivise individuals to act in the public interest. Attaching incentives in education may drive intrinsically motivated teachers to increase their effort and provide a higher standard of assistance to students. Accountability of teachers in less than ideal circumstances, where individual skill levels or organisational constraints exist, or whom are solely interested in seeking financial gain, may generate unintended outcomes as these individuals may elect to operate outside of or circumvent the system. “The difficulty with low performing teachers is that it is for them, in particular, that governments set standards and accountability stakes, but it is them, in particular, that have the least understanding of how to shift towards the standards” leaving them, at best, to comply superficially to external demands (Shalem, 2003, p.42). To the extent that school personnel are not motivated by stakes, the effort, instruction and quality of education provided will not improve and higher levels of performance will not result. This therefore means that the unfortunate unintended consequences of such accountability systems may be more prevalent than expected and the effect of these may be costly to students.

2.4 Gaming the System

2.4.1 What is Gaming Behaviour?

For those not motivated by the incentives and sanctions outlined in accountability programmes, high-stakes may result in unanticipated side effects as the system is “likely to induce behaviour distortions along other dimensions as agents seek to game the rules” (Jacob & Levitt, 2003, p.843). Whilst these stringent incentives and sometimes ruthless sanctions shock able teachers and other stakeholders into actively raising the quality of education they provide, those who are inadequate fear heavy sanctions and seek underhand methods of bolstering their performance. This is referred to as gaming the system or gaming behaviour. Holstrom and Milgrom (1991) find that accountability based on objective criteria will lead agents to focus on the most easily observable aspects of a multi-dimensional task. There is also a common belief that the performance measures associated with the accountability programme are flawed allowing agents to improve reported results without making actual progress on performance (Cullen & Reback, 2006). This suggests that those who circumvent the system may not be detected under the guise of improved student performance. This will lead to little actual improvement in the quality of education students’ experience.

2.4.2 Strategic Forms of Gaming Behaviour

There are a number of ways in which teachers and academically bankrupt schools are able to game the system in order to meet performance benchmarks or reflect adequate improvement. These methods include teaching to the test, reshaping the test pool by suggesting students be held back or pushed out of a grade thereby manipulating dropout rates, excluding academically poor students from the assessments and classifying certain students as disabled or limited-English (and therefore ineligible) These intentional acts reflect the manner in which assessment induces changes in teacher and sometimes student behaviour.

Teaching to the test

“There is increasing evidence that teachers’ ability to generate student learning is highly variable” (Bruns *et al.*, 2011, p.142). In light of this and under threat of sanctions, at risk teachers may focus intently on high-stakes subjects, ignore critical aspects of learning that are not explicitly tested and neglect those students who have no bearing on the accountability grade in order to raise performance. This is known as teaching to the test. As the standards of content are detailed and complex, “the assessments thus become a primary vehicle for communicating what the

standards really mean, and provide a strong signal to teachers and schools about what they should be teaching and what students should be learning” (Fuhrman & Elmore, 2004, p.142). Through repeated exposure to the content and style of national assessments, teachers are able to focus their instruction on the narrow range of the curriculum which is bound to be tested. This effectively allows teachers to reallocate their time towards helping students memorise tested content. Priority may also be given to only those specific subjects that are tested which usually include English and Mathematics to the detriment of subjects such as Science, History, Biology and other content areas (Linn, 2004). In addition to narrowing the curriculum, teachers are also able to coach students according to the pedagogical approach represented by the test. If national assessments make use of multiple choice questions that stress memorisation and routine computation rather than a deep understanding of the material and problem solving skills then the latter are likely to be ignored in the classroom. This is confirmed by Jacob (2005) who finds that the introduction of high-stakes testing in Chicago led to sharp increases in reading and math scores which were driven primarily by increases in test-specific skills and a substitution away from non-tested subjects such as science and social studies. Teachers have also been found to dedicate their attention and resources on those students at the margin who could affect the overall school grade, “while not necessarily improving the performance of other students” (Cullen & Reback, 2006, p.2).

Reshaping the test pool – holding back & increased punishment

Another method of gaming the system in order to achieve perceived improved performance involves reshaping the test pool. There is significant evidence on the number of ways in which schools and their administrators have been able to achieve this. Figlio and Getzler (2002) as well as Jacob (2005) find that the introduction of high-stakes testing is associated with a significantly higher rate of disability classification. By strategically placing at-risk students in disability or limited English programmes these students, under some accountability policies, may be exempt from the assessments or may benefit from testing accommodations reserved for students with a legitimate disability. Testing accommodations such as extra time, a quiet venue and a limited curriculum could improve any student’s performance, falsely raising the proficiency percentage of disabled students and the proficiency percentage of mainstream students within the school. Figlio (2006, p.837) further finds that schools assign harsher punishments to poorly performing students during the testing window, thereby using “discipline for misbehaviour as a tool to bolster aggregate performance.” This action by schools could put these students in greater risk as Jacob and Lefgren (2003) show that children who are home from school are more likely to engage in criminal activity. High-stakes testing is also associated with increased grade retention.

In Texas Cullen and Reback (2006) find that schools actively retain or exempt those students that could harm the school grade, disproportionately targeting low-performing Black and Hispanic students. Additionally the ChiPS study by Jacob (2005) reveals that among low-achieving students in the bottom schools, grade retention rates increased by up to 180 percent. These measures enable schools to temporarily circumvent accountability systems by superficially meeting proficiency targets.

Exclusion

Reshaping the test pool temporarily displaces academically poor students in order to meet benchmarks of adequate performance or progress. Exclusion, however, is permanent and whilst it may be beneficial to the school, it poses serious consequences for a student's potential future. In some instances, "it is actually to the school's advantage to drop slow learners and borderline students from the school, because they are usually poor test-takers" (Figlio & Getzler, 2002, p.2). The structure of accountability programmes could be said to incentivise schools to exclude or push out weak students as dropout rates account for a fraction of the accountability index compared to cognitive measures (Elmore, Abelman & Fuhrmann, 1996). A number of studies report this behaviour as low-scoring students are excluded from admission to "open enrolment" schools, encouraged to transfer to GED programs (for individuals who have left high school without a diploma) or asked to drop out of school altogether (Booher-Jennings, 2006; Darling-Hammond, 1991; Haney, 2000; Heilig & Darling-Hammond, 2008). Consequently even when students are not forcibly asked to leave schools they may do so voluntarily as schooling is perceived as confusing and pointless due to the degradation of the curriculum, zero tolerance towards absence and increased grade retention (McNeil *et al.*, 2008). In some cases, as in the state of Texas, this has resulted in real graduation rates of only 33 percent (McNeil *et al.*, 2008). Whilst the exclusionary effect of accountability incentives may be negated in future with more sophisticated built in incentives, the damage to date is alarming and highly costly to students.

2.4.3 Unintentional Strategic Behavior

A broader number of unintended side effects on school systems and students arise due to increased accountability including but not limited to content manipulation, myopia, complacency and indicator or grade fixation.

Content Manipulation

New accountability systems may unintentionally influence the content of the education syllabus thereby forgoing a broader education which focusses on imparting knowledge as well as ideals,

ethics and values for a more narrow view of subject material (de Wolf & Janssens, 2007). There is a significant amount of growing up that is done whilst a student is at school, but not necessarily engaged in learning. Under the *in loco parentis* role of the teacher, classroom dynamics, teacher and peer interactions and general discussion all contribute toward imparting ethics, manners, culture, and life lessons to students. This broader education deteriorates under accountability as teachers, particularly in academically bankrupt schools, are pressured to forgo the personal and intrinsic aspect of teaching in order to accommodate and prepare for national assessments.

Myopia

This subtle manipulation of standards may result in myopia where schools focus on short-term solutions at the expense of long-term policy reform (Smith, 2002). This is evident through teacher and school efforts to reach accountability targets and avoid sanctions rather than building and entrenching sound educational practices to improve student knowledge and understanding. Myopia underlies all gaming behaviour as teachers and at risk schools attempt to meet accountability standards falsely rather than pursuing a true account of affairs which could highlight areas of concern and draw assistance with regards to resources, consultants and the implementation of effective school systems. This is particularly disconcerting as it “takes about three years to achieve successful change in student performance in an elementary school [and] about six years to do so in a secondary school,” entrenching these bad practices (Fullan, 2002, p.1).

Complacency or Ossification

Accountability may also sanction complacency or ossification among well graded schools stifling innovation and experimentation in order to remain in the middle bracket (de Wolf & Janssens, 2007). This is seen at the classroom level where teachers place their focus on those marginal students that could affect the accountability grade and neglect the development of infra-marginal students (those well below or above the accountability grade) (Jacob, 2005). At the managerial or principal level, the fruit of investment in innovation or the experimentation of new practices is risky and may only reflect in future outcomes. It is therefore safer to focus on tried and tested policies and sustain moderate targets as exceptional performance raises expectations, “and the manager looking for a quiet life may instead spurn the short-lived plaudits arising from a beneficial innovation, and opt for chronic mediocrity” (Smith, 2002, p.236). Through accountability systems the incentive to solve the challenges facing the education system is engulfed by the incentive to meet benchmarks of adequate academic performance.

Indicator Fixation

High-stakes testing may also distort students' perspectives on their schooling experience making the "quest for an education a quest for grades and, by extension, for the more intrinsic rewards for which grades can be exchanged" (Aiken, 1991, p.725). As most accountability systems are aimed towards primary and secondary schooling, an excessive emphasis on examination performance may foster an unhealthy outlook on academics. According to Akerlof and Kranton (2002, p.1167) "a student's primary motivation is his or her identity," and a continual focus on test results may begin to define a student's perceptions of oneself. For a struggling student in a competitive environment this may lead to low self-esteem which is reported as one of the contributing features of an individual's propensity to cheat (McCabe *et al.*, 2001).

2.5 Intentional Cheating and Rotten Apples

Perhaps the worst potential outcome of school accountability programmes is that of outright cheating. The stress and strain of a new education programme and focus on test scores and grades may induce students, teachers and administrators to cheat in order to achieve the desired result. This is the study area of Jacob and Levitt (2003) in their paper, "Rotten Apples". Jacob and Levitt explore teacher and administrator cheating, referred to as 'teacher cheating', in Chicago public schools just prior to the implementation of the No Child Left Behind Act. Whilst this Bill had not yet been passed, the state of Chicago was required to annually test elementary students, rate schools based on their performance in these tests and intervene in schools that had not made sufficient improvement (Jacob & Levitt, 2003). The authors use the individual question answers given by students in grades three through eight who took the Iowa Test of Basic Skills (ITBS) between the years 1993 and 2000 to estimate the prevalence of teacher cheating. Teacher cheating encompasses a broad range of illicit activities such as changing student responses on answer sheets, announcing the correct answers in the test venue, and teaching students based on knowledge of the questions in the test (Jacob & Levitt, 2003). This is a very new analysis of cheating as most previous studies focus on student copying through specific cheating events, retesting suspicious performance, analysing erasure patterns and error similarity analysis (Aiken, 1991; Angoff, 1974; Bellezza & Bellezza, 1989; Holland, 1996; Wollack, 1997).

Jacob and Levitt design a new approach that uses two types of indicators namely unexpected test score fluctuations and unusual patterns of answers for students within a classroom. The presence of teacher cheating increases unexpected test score fluctuations as these students will likely experience unusually large gains in test scores one year followed by very small gains and even

possible declines the following year when they are exposed to a different teacher. Teacher cheating may also leave tell-tale signs of unusual patterns such as blocks of identical answers, unusual patterns of correlation across student answers or unusual response patterns on a student's exam. An example of such would be an academically poor student answering all the difficult questions towards the end of a test correctly whilst getting many of the easy answers incorrect. These two indicators will be very weakly correlated in classrooms which do not experience cheating, but highly correlated where cheating is present thus forming the underlying identification strategy of the paper. The authors find that cheating is prevalent in 4 to 5 percent of the classes in the sample with this most likely being a cautious estimate as the approach is shown to only partially detect illicit behaviour. In addition cheating is by nature a secretive endeavour and there are many other forms of cheating such as allowing students extra time that may not be noted by this identification. As confirmation Jacob and Levitt then run a simulation shuffling the students in order to demonstrate that there is nothing mechanical in their identification strategy that confounds the results.

2.6 The Case for South Africa

Based on the approach outlined by Jacob and Levitt (2003) and further work by Gustafsson (2014), this paper aims to discover the prevalence of cheating in the South African context. There are a number of reasons one would expect to discover cheating in the education system through the Annual National Assessments. Some of these contributing factors include corruption, lacking resources, and a global cheating epidemic.

Corruption

South Africa has a growing embedded culture of corruption and illicit behaviour which is especially common in public sectors. In 2014 South Africa was ranked 67th on the Corruption Perception Index (CPI) with a score of 44. The CPI measures perceived levels of public sector corruption worldwide, from 0 'completely corrupt' to 100 'completely clean', with a poor score likely reflecting "widespread bribery, lack of punishment for corruption and public institutions that don't respond to citizens' needs" (Transparency International, 2014, p.1). Fittingly the Auditor General, Kimi Makwetu, recently reported irregular expenditure of R25.7 billion for the 2014-2015 year with education, health, and public works sectors accounting for 45 percent of these irregularities (Auditor General South Africa, 2015). The key driver of these findings is uncompetitive and corrupt procurement processes which were found in 31 percent of the audits observed (Auditor General South Africa, 2015).

Deeper cases of corruption have been discovered in the education sector recently through the “Report of the Ministerial Task Team appointed by Minister Angie Motshekga to investigate Allegations into the selling of posts of educators by members of Teachers Unions and Departmental Officials in Provincial Education Departments” (Volmink *et al*, 2016). The task team investigated 81 cases of suspected corrupt appointments and found that these point to “widespread practices of improper and unfair influence affecting the outcomes of the appointment of educators” (Volmink *et al*, 2016, p.16). Overall these cases suggest that there is corruption and collusion among all levels of officials leading to a system of nepotism (Volmink *et al*, 2016, p.16). The degree of corruption was found to vary between provinces depending on the balance of power between the South African Democratic Teachers Union (SADTU) and other Unions, and the Department of Basic Education in that province. Hence, “where Districts work strictly according to regulated procedures and where their managerial and administrative staff members are persistent and consistent in carrying out their duties in accordance with a coherent system, the Teacher Unions in those areas are held in check and procedures and decisions are led by the department” (Volmink *et al*, 2016, p.18). Effectively this is only the case in one-third of South Africa’s provinces with evident corruption in the North West Province, Eastern Cape, KwaZulu-Natal and Limpopo and suspected corruption in Gauteng and Mpumalanga (Volmink *et al*, 2016).

Lacking Resources

Furthermore, job insecurity and a lack of resources and effective monitoring leave South Africa unable to adequately supervise instruction and assessment. Under-qualified or unqualified teachers, especially in rural areas, teach ill-disciplined, overcrowded classrooms with insufficient resources and inadequate infrastructure (Prew, 2009). “In many instances the South African Government fails not only to provide facilities such as libraries and laboratory material, but also crucial learning materials such as books” (Modisaotsile, 2012, p.4). Under these conditions, teachers are incentivised to illicitly help students through their assessments so as to not reflect poorly on their effort. On a government level, failure of appropriate inspection and monitoring allows teacher and school cheating to go unnoticed, increasing the acceptability of the practise and further degrading the quality of education.

Global Cheating

In addition to the culture of corruption ever growing in South Africa and the lack of resources, cheating among school and university students seems to have become a global phenomenon. The USA is currently struggling through a cheating ‘crisis’ at schools with 51 percent of students

having admitted to cheating on at least one exam in a 2012 survey of 23 000 high school students run by the Josephson Institute for Ethics (Josephson Institute, 2012). Further evidence of this behaviour is reported in work by McCabe *et al.* (2001), Galloway (2012), Nichols *et al.* (2007) and many others. This epidemic is not limited to the USA with similar experiences around the globe as students from varying countries such as Brazil, Thailand, Denmark, New Zealand, Argentina and Singapore, to name a few, believe that the “cheating culture [is] here” (Crittenden *et al.*, 2009, p.337). “Cheating has [therefore] become a pervasive problem in high school and college classrooms, and schools face a culture where the behaviour has become a normalised, accepted mechanism for gaining advantage” (Galloway, 2012, p.379).

South Africa is not immune to these issues with a hearing recently being held on incidents of mass cheating in the 2014 matric examinations. Education spokesperson Elijah Mhlanga states that “more than 5 300 candidates from all provinces are currently being investigated for various alleged irregularities, including mass copying arising from the 2014 Grade 12 final examinations” (Mhlanga, 2015). Cheating seems to be highly prevalent in KwaZulu-Natal and the Eastern Cape with 39 and 19 exam centres respectively being investigated on these counts (Mhlanga, 2015). In this context, an investigation into the prevalence of cheating in South African schools is certainly relevant with initial results suggesting a substantial number of particularly high levels of ‘strange patterns’ in the ANA data.

Chapter 3: The Data

3.1 Introduction to the ANA Data

This paper uses data from just over 800 public schools reflecting performance at the item level in the 2013 Verification Annual National Assessments carried out in South Africa. The ANA were introduced in 2011 and have been affirmed by government as a “significant learner achievement indicator of educational quality in the South African schooling system” (Department of Basic Education, 2013, p.1). The ANA were initiated in South Africa to reveal and address national educational realities within the historical and socio-economic contexts of the country. Prior to the ANA, the Matric examination was the only standardised method of evaluating the education system. The story outlined by the matric results, however, is not particularly informative given 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” (Spaull, 2013, p.3). As such, the ANA focus on assessing Mathematics and Language competencies in grades one to six and grade nine across all public schools, including special schools and state-funded independent primary schools, in the country. This allows for more rigorous monitoring and is intended to highlight those problems that hide lower down in the educational system.

In the 2013 phase almost seven million learners from more than 24 000 schools participated in these assessments (Department of Basic Education, 2013). Whilst the ANA are not attached to any consequences based on school performance, the Department of Basic Education (DBE) has high hopes for the programme stating that the “key outputs are to ensure high quality of teaching and learning, improved literacy and numeracy at schools, better National Senior Certificate examination performance as well as expanding early childhood development” (2013, p.1). The ANA programme is young, vast and has expanded quickly leading to questions of the reliability and validity of the data. As such the assessments of a representative sample of schools at the key stages of grades 3, 6 and 9 were conducted, monitored, marked, captured and processed by independent service providers, Deloitte and Nexia SAB&T. The data from this sample is called Verification ANA (V-ANA) and was used to confirm the reliability of the data emanating from the universal test results (Department of Basic Education, 2013). Of the 24 000 schools that participated in 2013, 2 164 schools were selected for the verification process. Designing, implementing and capturing assessments at this level is time consuming and costly and it is therefore important that the data be accurate and useable.

3.2 Completeness of the Data

The 2013 service providers intended to capture item-level data from around 45 schools per province, grade and subject; representing a subsample of the Verification ANA data. As evident in Table 1 below, this was generally achieved with the exception of grade 6 and 9 Home Language. A few schools in the data were excluded due to problems experienced when merging datasets in order to attain total school enrolment, resulting in an accurate representation of the V-ANA schools in this analysis. Overall the completeness of the data is good with close to the targeted number of schools per province, grade and subject having the required data. In addition there were almost no item-level responses missing across all grades, allowing for a complete analysis of strange patterns at the item-level.

The Universal ANA dataset, however, reports a number of data complications, the largest problem being that “around 15% of learners who should be reflected in the data are not” (Vilakazi & Gustafsson, 2014, p.1). The level of incompleteness varies drastically across districts and duplicate learners seem to remain even after duplicate records in terms of the ‘unique’ learner identifier have been removed. In addition, the master list of schools held by the DBE confirmed that the universal ANA dataset misreported 100 schools as public schools when they are in fact independent (Vilakazi & Gustafsson, 2014, p.1). These anomalies potentially show the introduction of gaming behaviour and detail the data integrity issues that need to be focussed on in future. These issues validate the use of the Verification-ANA data in this analysis.

The following table reflects the national statistics relating to the size and shape of the item-level V-ANA data under review. The structure and number of the items or questions asked varies across grade and subject.

Table 1: Details on Items per Test

Assessment	Schools	Mean Schools per Province	Mean Learners per School	Items	Items not 0-1	Details on items not 0-1 (e.g. 11:2 means 11 items had a 0-1-2 response structure)
Grade 3 Mathematics	392	43,6	24.0	34	6	6:2
Grade 3 Home Language	396	44,0	24.1	27	9	8:2, 1:6
Grade 6 Mathematics	387	43,0	23.6	50	18	11:2, 5:3, 1:4
Grade 6 Home Language	134	14,9	22.8	50	9	5:2, 3:3, 1:15
Grade 9 Mathematics	394	43,8	25.1	62	38	17:2, 7:3, 9:4, 5:5
Grade 9 Home Language	108	12,0	24.5	39	17	14:2, 1:5, 1:10, 1:20

Source: V-ANA (2013)

3.3 Values on Class within Grade

Due to the mechanisms employed by Jacob and Levitt, it is useful to know which students in the same grade are also in the same class. This information is important as teachers are likely to only have influence over their class and students are likely to only be able to copy off those sitting in the same classroom. The 2013 ANA data does include class information but inconsistencies in the way this was reported and captured has led to a third of the schools not having this information available. In preliminary work by Gustafsson (2014), analysis was conducted using two different methods in order to gauge the effect this variable has on the results. In one approach, class data was used where it seemed trustworthy and in the other class data was ignored altogether, as if all learners in the same grade were in the same class. The two methods produced very similar indicator values suggesting that the class data problem does not have much bearing on the results. This paper will henceforth discard class information but in future, attempts should be made to ensure this variable is consistently reported.

Chapter 4: Methodology

4.1 The Construction of Cheating Indicators

There are a few distinctions to take into account when applying Jacob and Levitt's cheating indicators to the South African ANA data. With the Iowa Test of Basic Skills (ITBS) data, Jacob and Levitt are able to follow the same students through their grades over a three year period. This contributes towards the certainty of their findings and allows for comparisons across years. The V-ANA data in this analysis offers more of a snapshot of the educational system at one point in time. Whilst these assessments have been carried out each year, no effort has been made to ensure repeated sampling of the same students or psychometrically-comparable test instruments. The design of the V-ANA data therefore leads to a lack of comparability across years and restricts the application of the Jacob and Levitt measures.

A further distinction concerns the differing format of the annual tests between the two programmes. The ITBS is a multiple-choice assessment where students can answer "A," "B," "C" or "D." Jacob and Levitt's analysis is therefore concerned with the specific responses students choose, and the unusualness or suspiciousness of those answers given a host of background characteristics. The ANA comprise varying forms of questions, not limited to multiple-choice. The marks allocated to questions differ according to difficulty and structure. In addition, the data reflects only the individual score obtained by the student per item. The focus of this analysis is therefore whether a student got a specific item right or wrong as opposed to whether they chose a particular answer. Some of the marks allocated to questions did not follow a simple 0-1 structure, as can be seen in Table 1, and were reconstructed to conform to this format. In these cases the raw score was rounded down to "0" or up to "1" depending on the percentage achieved. As an example; a score of 1 out of 5 would be rounded down to "0", whilst a score of 3 out of 4 would be rounded up to "1".

4.1.1 Indicator 1 (Unexpected Test Score Fluctuations)

Jacob and Levitt use the Iowa Test of Basic Skills data from Chicago public schools which tracks the same students over time through the grades from 1993 to 2000. As such their first indicator of cheating compares average classroom test gains for each classroom relative to all other classrooms in the same subject, grade and year. Test gains of specific classes can then be monitored across years in order to detect cheating. A classroom with a cheating teacher will

experience unusually large gains in test scores in the current year followed by much smaller gains the following year under a new teacher. These large fluctuations in student test scores over grades indicate that teacher cheating has occurred. As the ANA data does not track the same students each year, this indicator cannot be replicated.

4.1.2 Indicator 2 (Suspicious Answer Strings)

Measure 1: Suspicious Item Patterns over Time

Measure 1 looks at unusual identical blocks of item-level responses within a classroom. These answer patterns arise from the most basic form of teacher cheating, the correction of consecutive student responses following conclusion of the assessment. Using a multinomial logit model Jacob and Levitt predict the probability that each student chooses each possible answer for each question given the likelihood that similar students, based on background characteristics, will choose that particular response. The authors then consider all students and consecutive answers to find unusual blocks of item-level responses that are least likely to have arisen by chance. The more unlikely the unusual block of responses, the greater the probability that cheating has occurred. As Jacob and Levitt (2003, p.851) describe;

“Thus if ten very bright students in a class of thirty give the correct answers to the first five questions on the exam (typically the easier questions), the block of identical answers will not appear unusual. In contrast, if all fifteen students in a low-achieving classroom give the same correct answers to the last five questions on the exam (typically the harder questions), this would appear quite suspect.”

This measure, however, is not appropriate given the structure of the V-ANA data. As per Table 1, the V-ANA sample consists of approximately 24 students per school from roughly 45 schools per province. These students are sampled across the grade leaving very few observations per teacher with ill-reported class or academic achievement values. This differs drastically from the data used by Jacob and Levitt (2003) where the average school caters to 722 students of which 88 percent are tested and included in official reporting. This allows for an average of 127 student observations per grade, a sample approximately 5 times larger than that reflected in the V-ANA data. The per-grade sample size in the V-ANA data therefore makes suspicious “blocks” of responses highly unlikely.

Measure 2: Suspicious Blocks of Item Responses

This is the first measure that seems feasible given the ANA data.

“The second measure of suspicious answer strings involves the overall degree of correlation in student answers across the test. When a teacher changes answers on test forms, it presumably increases the uniformity of student test forms across students in the class” (Jacob & Levitt, 2003, p.851).

This measure captures more subtle versions of teacher cheating. As an example a teacher might correct different answers on different student’s tests or provide students with help or hints when asked. This form of cheating then does not create unusual blocks of item answers but will affect the distribution of student test scores and reflect a more homogenous class in terms of responses.

There are two outcomes of a student’s response; correct “1” or incorrect “0”. As such the following logit model is estimated to find the probability of a student’s binary response given the responses of other students tested within the same school:

$$\Pr(Y_{ils} = 1) = \frac{e^{\beta x_l}}{1 + e^{\beta x_l}}$$

where Y_{ils} indicates the response of learner l on item i in school s . The estimated coefficients are reported as log-odds and need to be converted into predicted probabilities in order to gain a meaningful interpretation of the results.

The predicted probability is calculated as:

$$p_{ils} = \frac{e^{\hat{\beta} x_l}}{1 + e^{\hat{\beta} x_l}} \text{ for the actual response given by the student.}$$

The residuals are then calculated for each choice the students could have made on each item:

$$\begin{aligned} e_{ils} &= 0 - \frac{e^{\hat{\beta} x_l}}{1 + e^{\hat{\beta} x_l}} \text{ when students response is "0"} \\ &= 1 - \frac{e^{\hat{\beta} x_l}}{1 + e^{\hat{\beta} x_l}} \text{ when students response is "1"} \end{aligned}$$

Two residuals are calculated per learner. In order to achieve a school level measure of the response to each item, the residuals for each response are summed across students within the same school:

$$e_{is} = \sum_l e_{ils}$$

The summed residuals are then squared and divided by the number of tested students within the school in order to accentuate outliers and normalise by school size.

$$v_{is} = \frac{\sum e_{is}^2}{n_s}$$

This statistic is analogous to the variance of student responses on each item within each school. M2, the second measure of suspicious strings is then just the school average of this variance statistic across all test items:

$$M2_s = \bar{v}_s = \frac{\sum_i v_{is}}{ni} \text{ where } ni \text{ is the number of items on the exam.}$$

This measure uses the overall degree of correlation in student answers across the test within a given school. A cheating school will therefore have a narrower band of student responses, a higher degree of correlation in student responses and will have a higher M2 value than a school that does not engage in cheating.

Measure 3: Suspicious within class variance (i.e. high variance in cross-question correlation of student responses)

The “third indicator of potential cheating is a high *variance* in the degree of correlation *across* questions” (Jacob & Levitt, 2003, p.851).

This measure uses the variance in the degree of correlation across questions as an indicator of cheating behaviour. If teachers correct student responses on certain questions or help students with specific questions then the within-school correlation on these questions will be particularly high. The correlation on other questions, which students answered independently, will however be typical. This leads to larger cross-question variances in student correlations in cheating schools.

This measure is then calculated as:

$$M3_s = \sigma_{v_s}^2 = \frac{\sum_i (v_{is} - \bar{v}_s)^2}{ni}$$

A school with a high variance in cross-question correlation will have a high M3 value and will be suspected of cheating.

Measure 4: Implausible school-level item averages relative to other academically comparable schools

The final measure, measure 4,

“compares the answers that students in one [school] give compared with the answers of other students in the system who take the identical test and get the exact same score” (Jacob & Levitt, 2003, p.852).

This measure looks at the distribution of a student’s responses within a school in comparison with the distribution of all other students’ responses who achieved the same aggregate score on the ANA. This essentially enables an assessment of where the students get their marks.

In calculating this measure let;

$$q_{ils} = 1 \text{ if learner } l \text{ in school } s \text{ answered item } i \text{ correctly}$$

$$= 0 \text{ otherwise}$$

$$A_l = \text{aggregate score of learner } l \text{ on the ANA}$$

$$nl_A = \text{no. of learners with an aggregate score of } A$$

The fraction of students answering each item correctly at each aggregate score level can then be defined as:

$$\bar{q}_i^A = \frac{\sum_{s \in \{z: A_z = A\}} q_{ils}}{nl_A}$$

The difference in student response patterns across all other students with the same aggregate score can then be calculated as:

$$Z_{ls} = \sum_i (q_{ils} - \bar{q}_i^A)^2$$

The school level deviation in student responses across aggregates is then calculated as:

$$M4_s = \sum_l (Z_{ls} - \bar{Z}^A)$$

This measure then determines how suspicious, at the school level, the aggregate score of a student is given their response pattern and the deviation from the average response pattern given by all other students with the same aggregate score.

A composite measure, JL3, is then constructed as:

$$JL3_{sb} = (pct_M2_{s,b})^2 + (pct_M3_{s,b})^2 + (pct_M4_{s,b})^2,$$

where *pct_M2* refers to the percentile of the M2 value for a specific school, to give an overall measure of the suspiciousness of school level answer strings in each subject.

4.2 Identification Strategy and Assumptions

Ideally, one would like to observe the suspicious measure values for guaranteed non-cheating schools and compare these to all other schools in the data to find an estimate of the prevalence of cheating. However, in the absence of an appropriate counterfactual, three assumptions help identify which schools are engaged in cheating behaviour.

1. Cheating increases the likelihood of suspicious response strings and therefore gives high values of these measures; M2, M3 and M4.
2. A high degree of correlation between the three cheating measures suggests that a school with a high composite measure, JL3, is a cheater school.
3. If cheating classrooms had not cheated they would look like non-cheating classrooms; they would have a relatively constant degree of cheating indicators and these measures would be closer to or equal to 0.

Assumption 1 holds due to the manner in which the cheating measures have been constructed as seen in Jacob and Levitt (2003). Accordingly, cheating classrooms will be concentrated in the upper end or tail of the distribution of suspicious answer strings.

This in turn should be confirmed by a higher JL3 measure, under assumption 2. The table below displays correlation matrices for Grade 3 and 6 in Mathematics and Language. As can be seen, the correlations between the different cheating measures are all positive resulting in high JL3

scores for “cheater” schools. The other flatter sections of the distribution which are associated with much lower suspicious answer scores should therefore include very few cheaters.

Table 2: Correlation Matrices of Cheating Measures

Grade 3 Mathematics						Grade 6 Mathematics					
	M2	M3	M4	JL3	Score		M2	M3	M4	JL3	Score
M2	1.00					M2	1.00				
M3	0.83	1.00				M3	0.88	1.00			
M4	0.43	0.34	1.00			M4	0.47	0.35	1.00		
JL3	0.47	0.38	0.49	1.00		JL3	0.38	0.22	0.70	1.00	
Score	0.12	0.14	0.03	0.08	1.00	Score	0.30	0.16	0.12	0.25	1.00
Grade 3 Home Language						Grade 6 First Additional Language					
	M2	M3	M4	JL3	Score		M2	M3	M4	JL3	Score
M2	1.00					M2	1.00				
M3	0.78	1.00				M3	0.77	1.00			
M4	0.26	0.27	1.00			M4	0.38	0.36	1.00		
JL3	0.42	0.35	0.29	1.00		JL3	0.36	0.28	0.39	1.00	
Score	0.43	0.26	0.11	0.20	1.00	Score	0.34	0.27	0.27	0.15	1.00

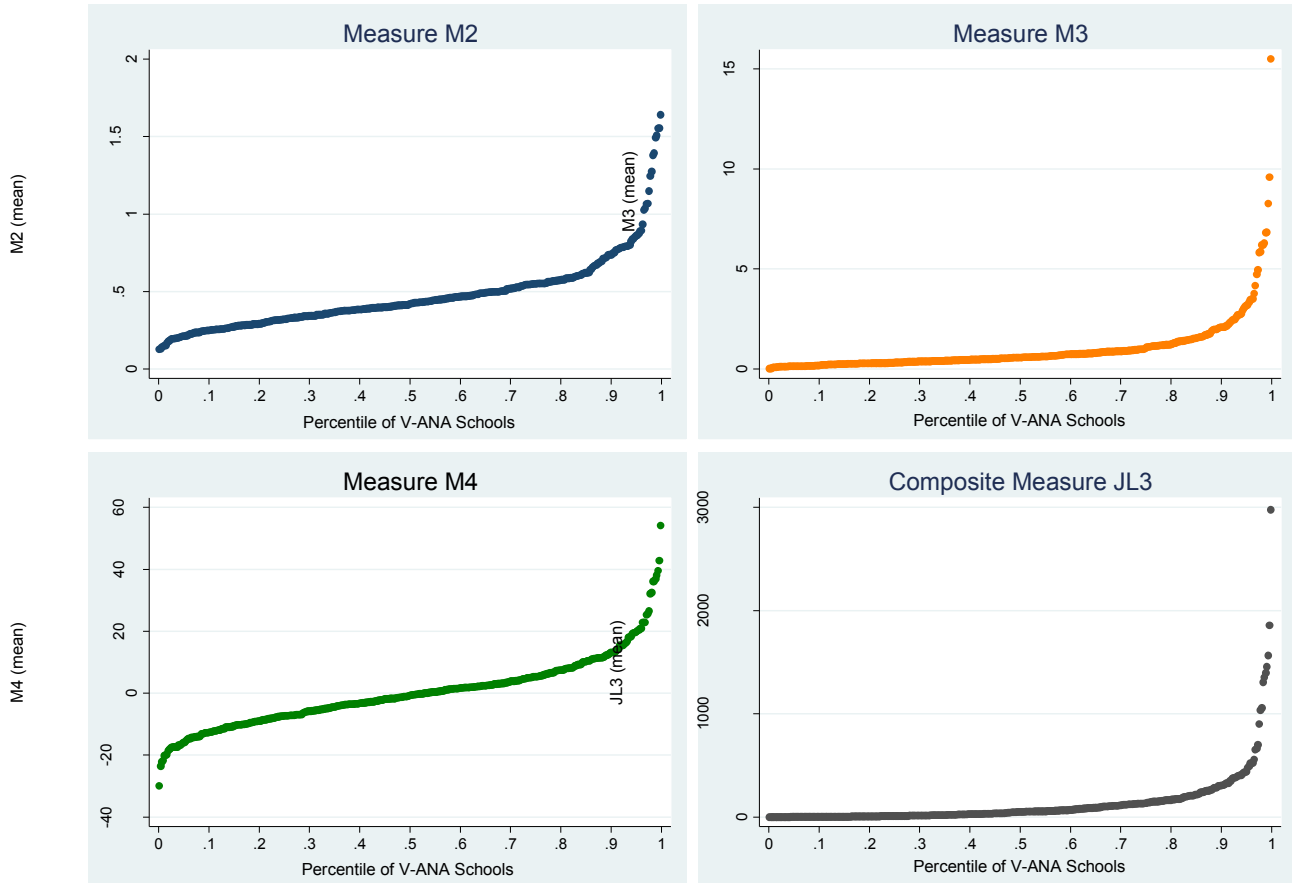
Source: V-ANA (2013)

Following assumption 3, schools in the lower and flatter section of the suspicious strings distributions can form the non-cheating counterfactual needed for analysis. The nature of this counterfactual is that it is situated in the bottom and middle range of the distribution. For analysis, however, this counterfactual is needed, but not observed, in the upper tail of the distribution where cheating occurs. Assuming assumption 3 holds, the observed counterfactual in the bottom and middle range of the distribution can be projected to the upper tail of the distribution to predict what it would have looked like in the absence of cheating. The distance between the predicted and observed distributions of suspicious answer strings then allows for an estimate of cheating to be calculated.

As an example, the following graphs show the percentile plots of the different and composite cheating measures for the Grade 3 V-ANA Mathematics assessment data. These graphs reflect the mean cheating scores and the associated percentage of schools who achieve that score. As seen below, the scores of the cheating measures increase rapidly towards the upper end of the distributions, particularly past the 85th and 90th percentiles. This indicates that cheating most likely is apparent in the schools past these points. In order to reduce the probability of labelling a non-cheating school in reality a “cheater” school in the analysis, the 90th percentile will be used

as a cautious cut-off point for suspicious measures in order to help quantify the extent of cheating in the ANA.

Figure 1: An Example of V-ANA Cheating Measures Percentile Plots



Source: V-ANA (2013)

4.3 Cheating Measures and Overall Levels of Performance

In order to ascertain that high measure values are driven by suspicious behaviour such as cheating, one needs to rule out all other possibilities. In this sense, it may be necessary to ascertain whether high measure values, especially M2 and M3, in certain provinces are driven by lower average school scores in these provinces due to the increased likelihood of blocks of incorrect answers and high cross-question variation. Under this hypothesis, schools with lower average school scores across all provinces would display higher M2 and M3 values compared to schools with higher average school scores. As seen in the previous correlation matrices, this hypothesis does not seem to hold according to the V-ANA data as is expected due to the manner in which the cheating measures were constructed. The correlation matrices show that across both subjects in Grades 3 and 6, higher M2 and M3 values do not seem to be associated

with lower scoring schools. In fact all correlation coefficients between all cheating measures and overall academic performance are positive indicating that on average, those schools with higher scores are the ones with the most suspicious data, high cheating measure values.

Overall, across all provinces, Grade 3 schools with suspicious data display better scores on average than schools with clean data. To illustrate the extent of this relationship, the average number of items correct for the 12 suspicious schools in the Mathematics assessment was 20.11 (these are schools with '1' scores for all four cheating measures) compared to an average score of 16.20 for the 288 clean or '0000' schools. The same pattern is observed in the Home Language assessment, the average number of items correct for the 8 suspicious schools was 15.46 (these are schools with '1' scores for all four cheating measures) compared to an average score of 10.69 for the 298 clean or '0000' schools.

The same relationship is observed with respect to Grade 6 schools whereby those with suspiciously high cheating indicator values on average perform better than those schools with lower suspicious string values. In the Mathematics assessment the average number of items correct for the 16 schools with excessively high suspicious measure values is 22.9 compared to an average score of 16.08 for the 294 clean schools. Similarly with respect to the assessment on First Additional Language on average the 7 "cheater" schools answered 25 items correctly compared to clean schools which achieved an overall score of 20.41 for the 210 clean schools. This gives an indication of the direction of causality: it would seem that on average schools that engage in suspicious or cheating behaviour display better overall scores, although in many cases these scores are not driven by their true academic achievement.

Chapter 5: Data Analysis and Findings

5.1 Summary Statistics: V-ANA Academic Performance

The following table shows the V-ANA average achievement level and pass rate for Grades 3, 6 and 9 in Mathematics and Home Language. Grade 3 Language and Mathematics is tested in all 11 official South African languages whilst Grade 6 and Grade 9 Language and Numeracy are written in English or Afrikaans. These figures corroborate the findings in the DBE's Report on the ANA of 2013. The DBE states that, "while the ANA results of 2013 point to an upward trend in most grades, we still have to strengthen our efforts towards realising the desired 60% threshold of learners mastering the minimum Language and Mathematics competencies by the end of Grades 3, 6 and 9" (Department of Basic Education, 2013, p.1).

Table 3: Summary Statistics on the Academic Performance of V-ANA Schools

Assessment	Mean	SD	Maximum Mark	No. Passed	Percent Passed
Grade 3 Mathematics	16.39	7.74	34	4620	49.03
Grade 3 Home Language	11.48	6.11	27	3634	38.00
Grade 6 Mathematics	16.62	10.49	50	1946	21.00
Grade 6 Home Language	32.11	10.87	50	2398	78.00
Grade 9 Mathematics	7.94	7.00	62	193	2.00
Grade 9 Home Language	18.75	7.26	39	1078	47.93

Source: V-ANA 2013

Generally the average attainment level for Language is higher and more consistent than that for Mathematics. This pattern continues when looking at pass rates, the percentage of students who attained more than 50 percent. Mathematics pass rates drop drastically between grades starting from a pass rate of almost 50 percent in grade 3 and culminating in a 2 percent pass rate in grade 9. This makes estimation for grade 9 mathematics particularly worrying due to the lack of variation in the data. Language pass rates remain more stable over grades with a great improvement between grade 3 and 6 up to a pass rate of 78 percent. Language pass rates do drop in grade 9 but not to the extent of Mathematics. Overall, the ANA results for grades 3 and 6 have improved since 2012.

5.2 Detailed Analysis of Grade 3

This section investigates and discusses the prevalence of cheating with respect to the Grade 3 V-ANA data. Results for both Mathematics and Home Language are reported and compared. Both Language and Mathematics assessments are conducted in all 11 official South African languages. Both the Language and Mathematics assessments vary in difficulty ranging from simple one mark answers or multiple choice questions to more difficult comprehensive or computational questions.

5.2.1 Summary Statistics for ANA Cheating Measures

The following table shows the summary statistics of cheating measures for Grade 3 Mathematics and Home Language. For Mathematics and Home Language 392 and 396 schools are discussed respectively. Schools presenting any cheating measure value above the 90th percentile were deemed to reflect high levels of suspicious responses and were labelled “cheater” or “noted” schools. Most of the cheating indicators reported below are lower for Home Language than for Mathematics suggesting that cheating occurs to a greater degree in higher-stakes subjects like Mathematics.

Table 4: Summary Statistics on Cheating Measures for Grade 3

Assessment	Mean	SD	Min	10th %	Median	90th %	Max	Noted Schools
Mathematics Grade 3								
M2 ANA	0.44	0.20	0.13	0.25	0.41	0.62	1.64	58
M3 ANA	0.84	1.25	0.02	0.14	0.50	1.60	15.48	54
M4 ANA	-0.20	10.96	-30.01	-12.68	-0.85	12.17	54.11	43
Home Language Grade 3								
M2 ANA	0.45	0.19	0.03	0.25	0.42	0.69	1.36	67
M3 ANA	0.59	0.72	0.01	0.11	0.37	1.37	10.16	33
M4 ANA	-0.05	8.15	-23.20	-9.91	-0.73	10.42	29.32	27

Source: V-ANA (2013)

M2 values are almost identical between Mathematics and Home Language while M3 and M4 values begin to differ. The average M3 value for Mathematics is 0.84 compared to an average M3 value for Home Language of 0.59. M4 values diverge further with mathematics M4 values on average 1.35 times the magnitude of those for Home Language. The M4 measure gives an overall measure of the implausibility of a school’s responses. Restricting the analysis to consider only this measure, just over 11 percent of schools display suspicious answer strings in Mathematics

compared to almost 10 percent for Home Language. Suspicious response patterns are therefore prevalent in both subjects; however, the suspicious measure values and 90th percentile cut-off points are lower for Home Language suggesting that cheating behaviour may be more obvious in higher stakes subjects like Mathematics which is consistent with international literature on high stakes testing.

The following table shows combinations of high measure values for Grade 3 Mathematics and Home Language. Schools exceeding the 90th percentile values for the individual cheating measures, or composite measure, were considered to have high measure values and were assigned a value of “1”. This table acts as a cursory check and provides a more conservative estimate of the prevalence of cheating in the ANA data. If one assumes that only those schools exhibiting high measure values across all cheating measures are engaged in suspicious behaviour then only 3.06 percent and 2.02 percent of schools can be labelled as “cheater schools” in Mathematics and Home Language respectively (using the 90th percentile threshold).

Table 5: Combinations of High Value Cheating Measures for Grade 3

Suspicious Answer String Measures				Noted Schools			
M2	M3	M4	JL3	Mathematics	(%)	Home Language	(%)
0	0	0	0	288	73.47	298	75.25
0	0	1	0	8	2.04	11	2.78
1	1	0	1	1	0.26	0	0
1	1	1	1	12	3.06	8	2.02
0	1	0	0	12	3.06	14	3.54
1	0	0	0	9	2.3	12	3.03
1	1	0	0	24	6.12	19	4.80
0	0	1	1	11	2.81	13	3.29
Other less frequent combinations				27	6.89	21	5.30
Total Schools				392		396	

Source: V-ANA (2013)

Using M4 values and high measure values across all cheating measures provides a conservative or lower bound estimate of the prevalence of cheating in the V-ANA data. According to these constructs, cheating occurs in between 3 percent and 11 percent of schools for Mathematics and 2 percent and 10 percent for Home Language in Grade 3. These estimates may in fact be greater considering the large standard deviations associated especially with the M4 measure. It is difficult to believe that suspicious or cheating behaviour may be organised by Grade 3 students which may point to the pressure teachers feel when they believe they are being scrutinised or rated.

These summary statistics indicate that even in lower primary schooling, teachers perceive the Annual National Assessments to be a reflection of their effort and respond as if the assessments involve high stakes.

5.2.2 Analysis of Cheating Measures by Province

The national perspective previously discussed gives an idea of the extent of cheating in South Africa within the ANA in Grade 3. It may also be interesting, however, to analyse the extent of strange patterns by province as provincial departments play a large role in determining the types of controls and monitoring that occur during testing. This will highlight those provinces that where cheating is more prevalent and require additional intervention or support to address these issues. Local policy of these provinces can then implement new structures to prevent or mitigate the spread of cheating in education. This is particularly important as, at the Grade 3 level, “6 of the 9 provinces have a failure rate of over 50 percent” (Deloitte & Nexia SAB&T, 2014, p.5).

The following series of tables reflect the extent of suspicious behaviour by province for the Grade 3 Mathematics and Home Language assessments. Table 6 shows the actual 90th percentile values of the cheating measures by province. A cursory glance at the data reveals that suspicious behaviour seem to be a problem specifically in the Eastern Cape but also in KwaZulu-Natal and Mpumalanga in Mathematics specifically and to a lesser extent in the Free State and KwaZulu-Natal in Home Language.

Table 6: 90th Percentile Cheating Measure Values by Province for Grade 3

Province	Mathematics				Home Language			
	M2	M3	M4	N	M2	M3	M4	N
EC	0,79	3,49	32,12	41	0,78	1,48	5,49	42
FS	0,59	1,24	11,16	41	0,79	1,64	12,32	45
GP	0,59	1,42	9,23	53	0,59	1,12	9,29	55
KZN	0,8	3,19	25,29	42	0,9	2,12	15,26	41
LP	0,77	2,08	13,13	42	0,6	0,75	8,92	40
MP	0,67	1,95	16,76	40	0,65	1,18	9,91	42
NC	0,5	0,83	10,05	44	0,58	0,89	8,72	44
NW	0,61	1,48	13,16	45	0,72	0,71	8,78	43
WC	0,53	0,9	3,52	44	0,66	1,15	10,41	44
National 90th Percentile	0,62	1,6	12,21	392	0,69	1,37	10,42	396

Source: V-ANA (2013)

Within Mathematics; the Eastern Cape, KwaZulu-Natal and Mpumalanga have exceptionally high values, particularly M4 values. The average Eastern Cape and KwaZulu-Natal values are double those of the national 90th percentile for M3 and M4 values, indicating an excessive amount of strange response patterns which points to suspicious behaviour. Mpumalanga's suspicious values do exceed the national 90th percentile values for Mathematics, but only marginally. With respect to Home Language, only KwaZulu-Natal and the Free State display particularly high values. These provinces exceed the national 90th percentile values across all suspicious string indicators.

The 90th percentile values reported above have been used to present tables 7 to 10, which show these results in a more accessible manner. Tables 7 and 8 present the number and percent of schools which achieved cheating measure values above the national 90th percentiles respectively whilst tables 9 and 10 report the number and percent of these noted schools with cheating measure values at least two standard deviations above the national 90th percentile values respectively, by province. These tables reiterate that the Eastern Cape and KwaZulu-Natal are the most suspicious in the Mathematics assessment whilst the Free State and KwaZulu-Natal show suspicious behaviour in the Home Language assessment.

With respect to Mathematics, the Eastern Cape and KwaZulu-Natal show a high number of noted schools in each cheating measure. This is further confirmed in table 10 where 5 percent and 2 percent of schools in these provinces respectively have cheating measures greater than two standard deviations above the national 90th percentiles. Over all the suspicious response measures, the Eastern Cape reports 6 noted schools (14.6 percent) whilst KwaZulu-Natal shows 4 noted schools (9.5 percent). Suspicious blocks of item responses (M2) are the most common form of strange response pattern in the Eastern Cape whilst high cross-question variance (M3) is most common in KwaZulu-Natal schools.

Suspicious behaviour does not seem to be as pronounced in the Home Language assessment. Whilst the data shows at least one noted school per cheating measure within each province, when looking at noted schools across all measures only the Free State and KwaZulu-Natal stand out. Both provinces report 3 noted schools (approximately 7 percent), however these provinces show different patterns of suspicious behaviour. The Free State reports strange response patterns compared to similar schools (M4) whilst KwaZulu-Natal shows suspicious blocks of item responses (M2). Whilst approximately 7 percent of schools in the Free State and KwaZulu-Natal show overall levels of suspicious behaviour, there are no schools with cheating measure

values two standard deviations above the national 90th percentiles. This reflects the fact that suspicious behaviour is more prominent in high-stake subjects.

Table 7: Number of Noted Schools by Province for Grade 3

Province	Mathematics					Home Language					
	M2	M3	M4	All	N	M2	M3	M4	All	N	
EC	15	13	10	6	41	8	6	1	0	42	
FS	3	2	4	1	41	6	7	8	3	45	
GP	0	0	0	0	53	3	8	4	0	55	
KZN	10	11	8	4	42	11	9	8	3	41	
LP	10	10	6	1	42	3	2	5	1	40	
MP	6	6	6	1	40	5	4	4	1	42	
NC	2	1	0	0	44	1	3	3	1	44	
NW	7	6	6	3	45	5	3	2	0	43	
WC	1	1	0	0	44	3	3	4	0	44	
Total	54	50	40	16	392	45	45	39	9	396	
National 90th Percentile Cut-offs Applicable:											
M2					0.62	M2					0.69
M3					1.60	M3					1.37
M4					12.21	M4					10.42

Source: V-ANA (2013)

Table 8: Percent of Noted Schools by Province for Grade 3

Province	Mathematics					Home Language					
	M2	M3	M4	All	N	M2	M3	M4	All	N	
EC	37	32	24	15	41	19	14	2	0	42	
FS	7	5	10	2	41	13	16	18	7	45	
GP	0	0	0	0	53	5	15	7	0	55	
KZN	24	26	19	10	42	27	22	20	7	41	
LP	24	24	14	2	42	8	5	13	3	40	
MP	15	15	15	3	40	12	10	10	2	42	
NC	5	2	0	0	44	2	7	7	2	44	
NW	16	13	13	7	45	12	7	5	0	43	
WC	2	2	0	0	44	7	7	9	0	44	
National 90th Percentile Cut-offs Applicable:											
M2					0.62	M2					0.69
M3					1.60	M3					1.37
M4					12.21	M4					10.42

Source: V-ANA (2013)

Table 9: Number of Schools at least 2 Standard Deviations above the National 90th Percentile for Grade 3

Province	Mathematics					Home Language				
	M2	M3	M4	All	N	M2	M3	M4	All	N
EC	3	4	2	2	41	1	2	0	0	42
FS	1	0	0	0	41	0	1	1	0	45
GP	0	1	0	0	53	0	1	0	0	55
KZN	2	1	3	1	42	3	2	1	0	41
LP	2	3	2	0	42	1	1	0	0	40
MP	1	1	0	0	40	0	0	0	0	42
NC	1	1	0	0	44	1	1	0	0	44
NW	3	2	0	0	45	0	0	0	0	43
WC	0	0	0	0	44	0	2	0	0	44
Total	13	13	7	3	392	6	10	2	0	396
National Cut-offs Applicable:										
M2	1.03				M2	1.07				
M3	4.11				M3	2.82				
M3	34.12				M4	26.73				

Source: V-ANA (2013)

Table 10: Percent of Schools at least 2 Standard Deviations above the National 90th Percentile for Grade 3

Province	Mathematics					Home Language				
	M2	M3	M4	All	N	M2	M3	M4	All	N
EC	7	10	5	5	41	2	5	0	0	42
FS	2	0	0	0	41	0	2	2	0	45
GP	0	2	0	0	53	0	2	0	0	55
KZN	5	2	7	2	42	7	5	2	0	41
LP	5	7	5	0	42	3	3	0	0	40
MP	3	3	0	0	40	0	0	0	0	42
NC	2	2	0	0	44	2	2	0	0	44
NW	7	4	0	0	45	0	0	0	0	43
WC	0	0	0	0	44	0	5	0	0	44
National Cut-offs Applicable:										
M2	1.03				M2	1.07				
M3	4.11				M3	2.82				
M3	34.12				M4	26.73				

Source: V-ANA (2013)

The above tables display sharp contrasts between those provinces that do display strange response patterns which hint at cheating and those that don't. In general, suspicious measures are noticeably higher for Mathematics than for Home Language which corroborates international

literature. Within Mathematics; the Eastern Cape and KwaZulu-Natal are the provinces showing the most pronounced suspicious behaviour with 14.6 percent and 9.5 percent of schools exceeding the national 90th percentiles across all measures respectively. With respect to Home Language, the Free State and KwaZulu-Natal report suspicious behaviour however not to the extent of that seen in the Mathematics assessment, approximately 7 percent of schools. The manner in which schools behave suspiciously varies significantly by province, subject and measure. More than a third of schools (37%) in the Eastern Cape and about a quarter (24%) of schools in schools Limpopo and KwaZulu-Natal have suspiciously high M2 and M3 values for the grade 3 Mathematics assessment compared to 2 percent in the Western Cape and 0 percent in Gauteng.

5.2.3 Patterns within Individual Schools

This paper has referred to the suspicious patterns of schools and thus far has discussed only the suspicious measure values associated with these schools. This section will explore individual suspicious schools and compare these to the average patterns displayed by similar clean schools. In this instance similar clean schools refer to those unsuspecting schools which on average achieved the same overall score on the assessment as the suspicious school in question. Comparing the patterns of individual schools allows an assessment on whether these abstract adapted Jacob and Levitt measures are meaningful when applied to the V-ANA data.

The following series of graphs depict some of the most suspicious schools by cheating measure in comparison to similar clean schools. The items referred to on the horizontal axis are sorted according to the values of the similar schools' average. The perceived more difficult items are therefore displayed on the left of the graphs whilst the easier items are displayed to the right. Non-suspicious school response patterns, those with low M2, M3, M4 and JL3 values, are expected to follow the similar clean schools average quite closely whilst highly suspicious school response patterns, those with high M2, M3, M4 and JL3 values, should diverge in varying degrees from the similar schools' average.

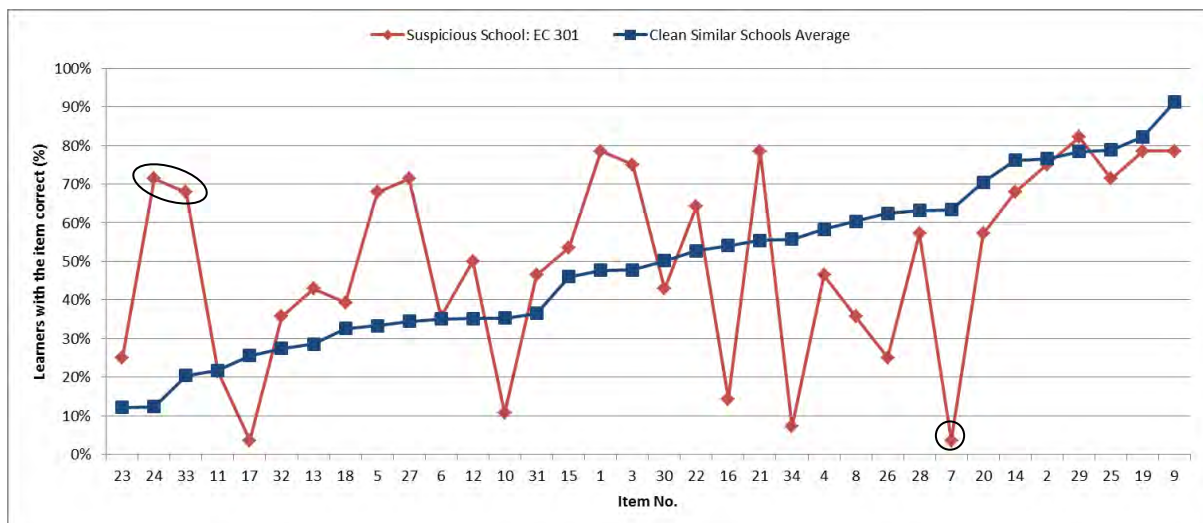
5.2.3.1 Mathematics

There are 12 schools in total that have excessively high suspicious measure values across all three cheating measures as well as the composite measure, JL3. The highest suspicious measure values for M2, M3, M4 and JL3 respectively are 1.55, 9.58, 54.11 and 2976.40. The average assessment score associated with these suspicious schools is 20.11 compared to an average score of 16.20 for the 288 clean schools. Of the 12 schools that are in the top ten percentiles across all cheating

measures, four are from KwaZulu-Natal, another four are from the Eastern Cape, three are from the North West province and one is from Gauteng.

Figure 2 below graphs the most suspicious school according to M4 and JL3 values, arguably the worst culprit. Test data was available for 28 learners in this Eastern Cape school (EC 301) who on average achieved a score of 16.8 out of 34. The similar schools' average is based on 21 schools who achieved an overall average score of between 16 and 17 out of 34. The suspicious school's response pattern per item diverges significantly from that of the similar schools' average. There are a number of instances where the school in question (EC 301) significantly outperformed the average on difficult items and drastically underperformed on some of the easier items.

Figure 2: Illustration of Strange School EC 301



Source: V-ANA (2013)

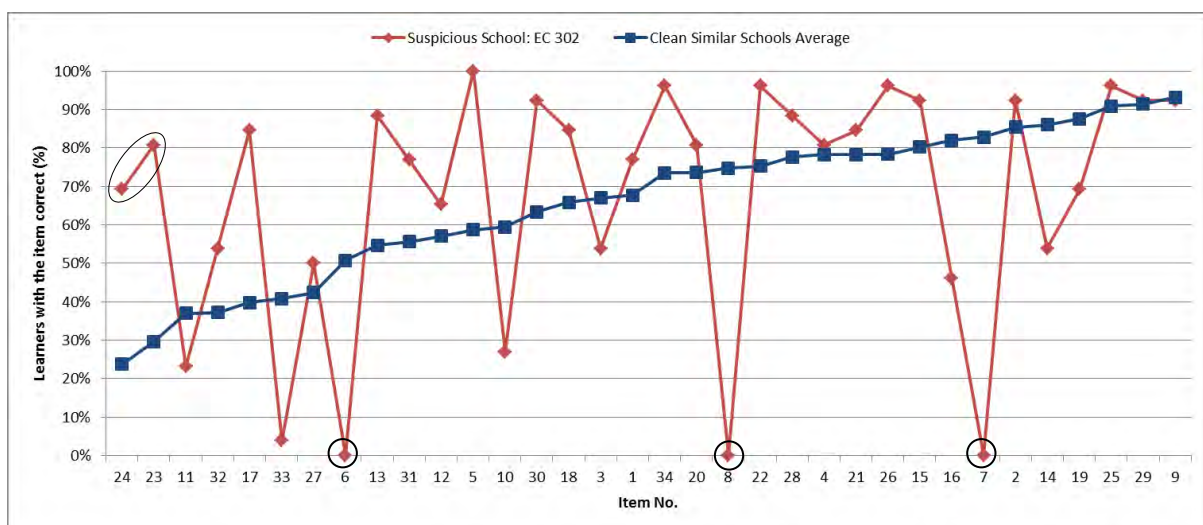
Items 24 and 33 reflect the most stark examples of where EC 301 significantly outperformed similar unsuspecting schools. These items were perceived to be very difficult according to the similar schools' average with only 12 percent and 20 percent of comparable learners answering correctly. It is therefore strange that approximately 70 percent of learners in EC 301 were able to answer both of these items correctly. In the other direction on item 7 towards the easier direction of the graph only 3 percent of EC1 learners answered the item correctly compared to an average of 61 percent of learners in similar clean schools.

Overall the suspicious school in question appears to display no general pattern but rather exceeds general expectations on more difficult questions of the assessment and underperforms on easier questions of the assessment leading to the overall highest M4 and JL3 values. This

school therefore gains and loses its marks in very unlikely sections of the assessment and on average diverges from similar clean schools by 20 percent.

The following figure displays the average response pattern from the school with the highest M2 value in comparison to similar schools. This school should therefore display blocks of suspicious responses. This Eastern Cape school (EC 302) on average achieved a score of 22.9 out of 34 based on the data available for 26 learners who took part in the assessment. The similar schools' average is based on 12 clean schools who achieved an overall average score of between 22 and 23 out of 34.

Figure 3: Illustration of Strange School EC 302



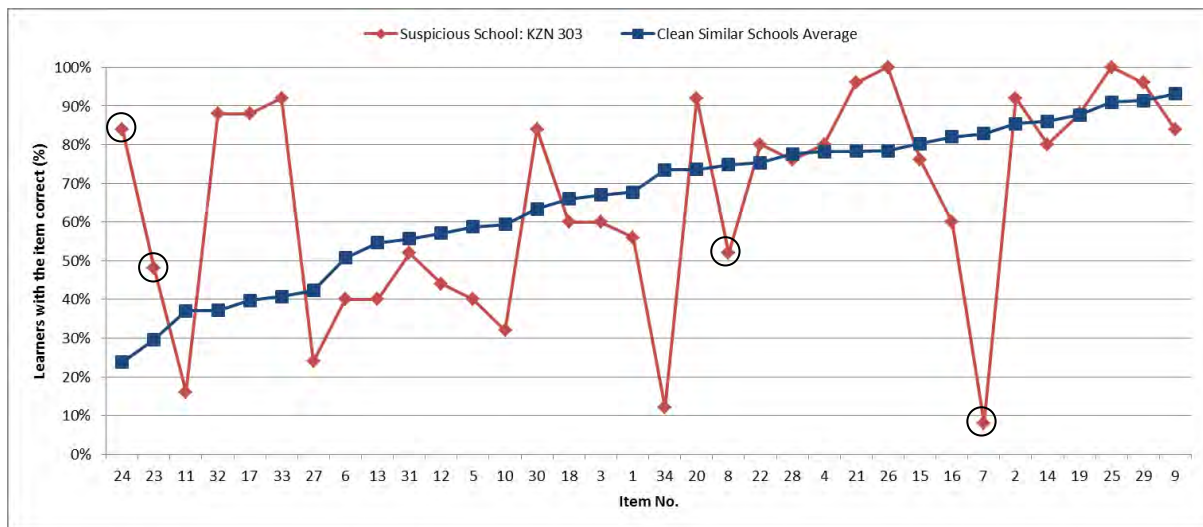
Source: V-ANA (2013)

This school shows two particularly obvious cases of questionable block responses as well as a tendency to outperform the average. Despite generally performing better than similar schools, not a single student in EC 302 answered items 6, 7 and 8 correctly. This is made more suspicious due to the varying degrees of difficulty of these items. Items 7 and 8 were perceived to be much easier than item 6 so one would expect that at least some students in EC 302 would answer these items correctly. Learners in clean comparable schools fared significantly better than EC 302 with approximately 45 percent, 72 percent and 70 percent of students answering these items correct respectively. Item 23 and 24 reflect another case of questionable student responses where the suspicious school in question outperformed similar schools by approximately 40 percent on consecutive questions deemed by students to be very difficult. The suspicious blocks of strings evident in this schools average response pattern drives the M2 value upwards.

In addition to suspect blocks of item responses, EC 302 shows particularly extreme results. Using a threshold of 95 percent of students answering a question correctly, the school in question displays 5 items fulfilling this criterion against 2.5 items for similar schools. Similarly EC 302 reports 4 items compared to 1 when analysing the number of items where less than 10 percent of students got the answer correct. These blocks of uniformed responses and extreme positive and negative average student responses points to some form of assessment manipulation.

Figure 4 below reveals the average response pattern from the school with the highest M3 value in comparison to similar clean schools. This school should therefore display suspiciously high cross-question variance. The average school response pattern for KZN 303 is calculated from the 25 students who took part in this test and who on average achieved a score of 22.2 out of 34. The similar schools' average is based on 12 schools which achieved an overall average score of between 22 and 23 out of 34.

Figure 4: Illustration of Strange School KZN 303



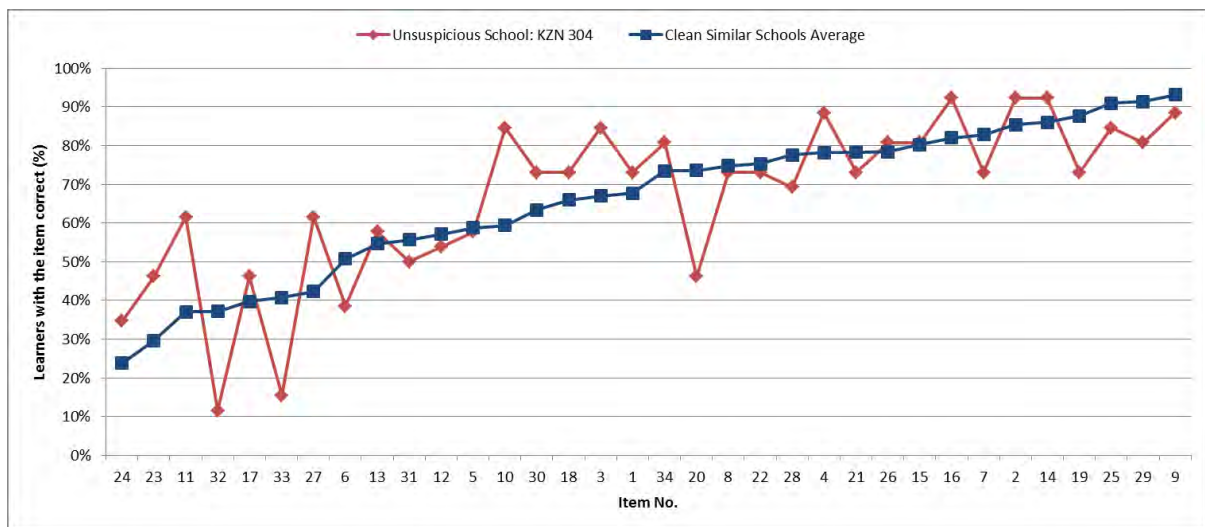
Source: V-ANA (2013)

The school in question, KZN 303, reflects large differences in the percentage of students answering consecutive questions correctly. This is particularly evident over items 7 and 8, and 23 and 24. This school faired much worse than comparable clean schools over items 7 and 8, which were deemed to be relatively easy. In addition to underperforming over these questions, there is a 40 percent increase in the number of students with these items correct between item 7 and item 8 in KZN 303. A similar occurrence is seen over items 23 and 24 which were perceived by similar students to be very difficult. In this instance the suspicious school performed

considerably better than similar clean schools. Just less than 50 percent and over 80 percent of students in the strange school answered items 23 and 24 correctly compared to similar clean schools where on average 40 percent and 34 percent of students answered these items correctly respectively. Overall student responses in KZN 303 seem to vary greatly across consecutive items and display no particular pattern.

The following school is shown to be clean, presenting very low values for the cheating measures created. There is test data on 26 students in this school who on average answered 22 out of 34 items correct on the assessment. The similar schools' average is based on 12 schools who achieved an overall average score between 22 and 23 out of 34. This school diverges from the average generally for questions considered more difficult by similar schools. In these instances the school in question fares worse than the average which is to be expected. On questions deemed slightly easier, this unsuspecting school has outperformed the average. Overall KZN 304 follows the average pattern of similar clean schools which is suggested by the low cheating measure values.

Figure 5: Illustration of Non-Strange School KZN 304



Source: V-ANA (2013)

It is interesting to note that many of the suspicious schools struggled or performed well over the same items. In general these schools performed better than similar clean schools on more difficult questions such as items 23, 24 and 33 and fared worse than similar clean schools on easier questions such as items 6, 7 and 8. Some of the items over which suspicious schools deviated most from the similar schools' average are displayed below.

Box 1: Grade 3 Mathematics Items on which Suspicious Schools Outperformed the Average

Item 23: Question 19

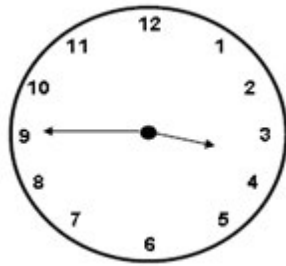
19) John, Jim and Joan each have 37 lollipops. How many lollipops do they have altogether?

Item 24: Question 20

20) There are 36 chocolates in a box. Each child gets 5 chocolates. How many children got chocolates? How many chocolates were left over?

Item 33: Question 26

26) Complete the sentence below:



The time on the analogue clock reads _____.

Source: ANA (2013), Department of Basic Education

Box 2: Grade 3 Mathematics Items over which Suspicious Schools Underperformed

Item 6: Question 6

6) Arrange 1 half, 1 quarter, 1 third, 1 fifth from the greatest to the smallest.

- A 1 quarter, 1 half, 1 fifth, 1 third
- B 1 half, 1 third, 1 quarter, 1 fifth
- C 1 third, 1 fifth, 1 quarter, 1 half
- D 1 fifth, 1 quarter, 1 third, 1 half

Item 7 & 8: Question 7.1 & 7.2

7) Write the name of the given 3D object and 2D shape.

7.1)



7.2)



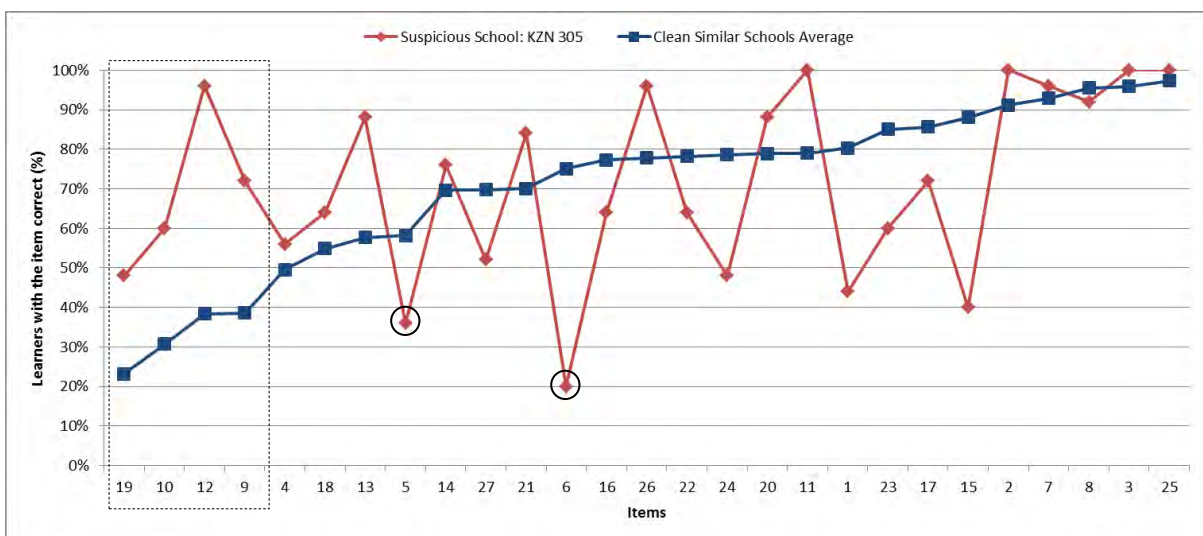
Source: ANA (2013), Department of Basic Education

5.2.4.2 Home Language

There are 8 schools in total that have excessively high values across all three cheating measures, M2, M3 and M4, as well as the composite measure, JL3. The highest suspicious measure values for these measures are 1.36, 10.16, 29.32, and 863.30. The average assessment score associated with these suspicious schools is 15.46 compared to an average score of 10.69 for the 298 clean schools. Of these 8 schools that are in the top ten percentiles across all cheating measures, three are from KwaZulu-Natal, another two are from the Free State, one is from Limpopo and another is from the Eastern Cape.

Figure 6 graphs the average item responses of the most suspicious school according to M4 and JL3 values. Student responses are available for the 25 learners who wrote the Home Language assessment and on average achieved a score of 19.2 out of 27 in this KwaZulu-Natal school, KZN 305. The similar schools' average is based on 12 schools who achieved an overall average score of between 18 and 21 out of 27. The suspicious school's responses show a general pattern of underperforming on easier questions and over-performing on more difficult questions compared to similar clean schools. Whilst this school reflects the worst M4 score out of the schools that are above the 90th percentile in all the cheating measures for Home Language, this school's M4 value is just over half that of the worst M4 value found in the most suspicious Mathematics school. This school's item response pattern therefore does not diverge as far away from the similar schools' average as was previously seen in the worst culprit for Mathematics.

Figure 6: Illustration of Strange School KZN 305



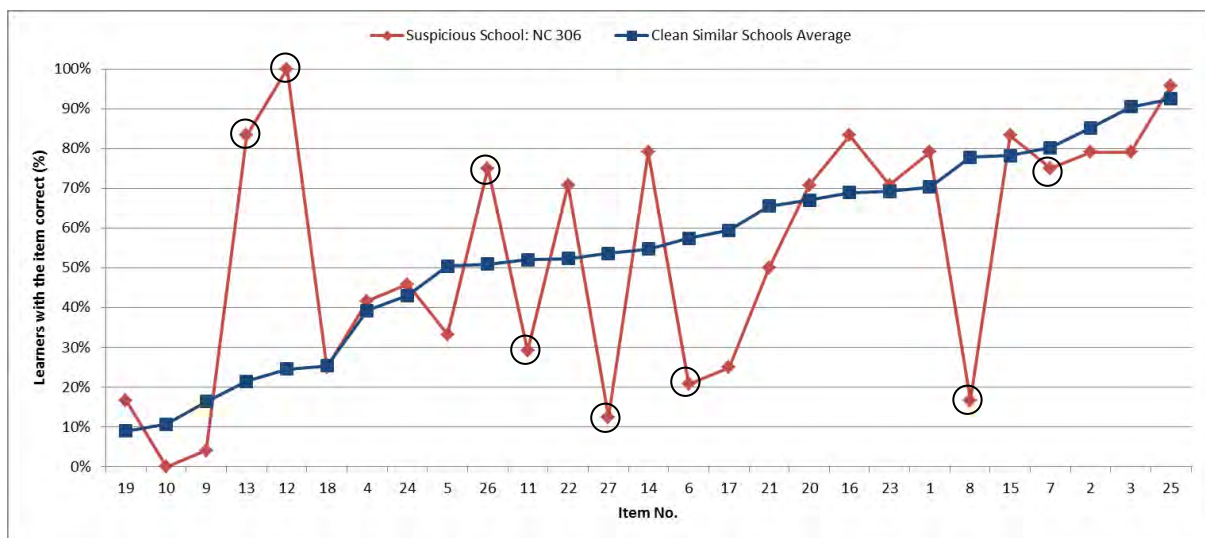
Source: V-ANA (2013)

There are a number of average student responses which suggest that learners in this school achieved their marks in a dubious manner. Items 5 and 6 were deemed relatively easy comprehension questions with 60 percent to 75 percent of comparable students in similar schools answering these consecutive questions correctly. KZN 305, however, performed considerably worse with an average of only 36 percent and 20 percent of students answering these respective items correctly. In the other direction this school consistently outperformed similar clean schools by approximately 36 percent over 4 of the most difficult items.

Overall students in this school gained most of their marks over the more difficult questions in the assessment, whilst losing marks on the easier section of the assessment. This school does display implausible school-level item averages; however, these are not as obvious as those seen in the Mathematics assessment.

The following school displays the worst M2 value and second worst M3 value out of the 8 suspicious schools. The average score achieved by the 24 learners in this school was 14.46 out of 27. The similar schools' average is based on 12 schools where students on average achieved score of between 14 and 15 items correct out of a potential 27. This Northern Cape school does show definite signs of suspicious behaviour in the form of suspicious blocks of student responses and high cross-question variance.

Figure 7: Illustration of Strange School NC 306



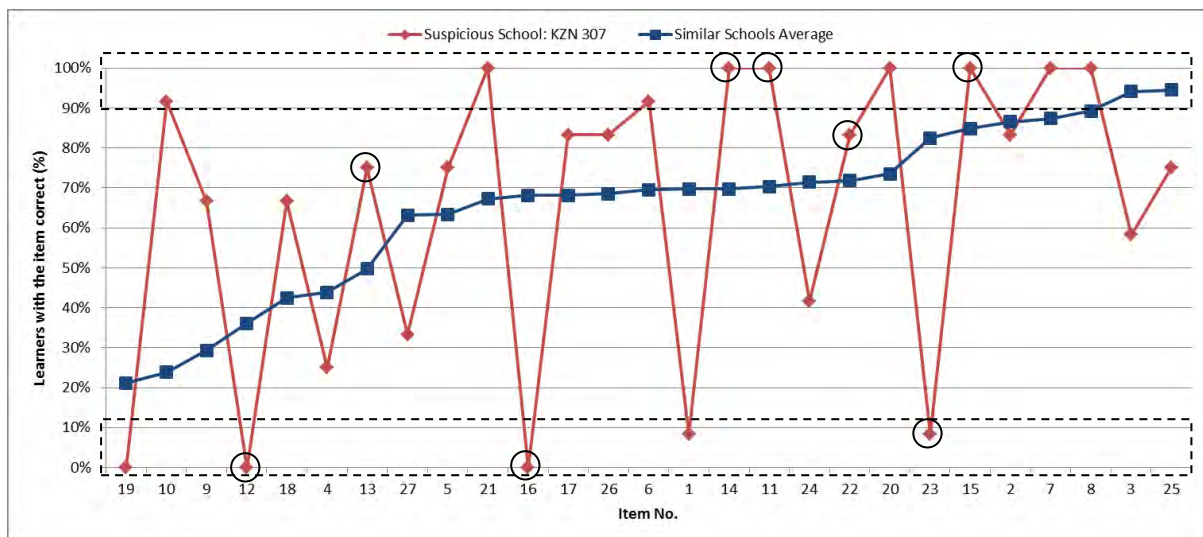
Source: V-ANA (2013)

Over items 12 and 13, NC 306 on average outperformed similar schools by nearly 70 percent. In addition item 11 was the first and considerably much easier of the three part question, however, only 30 percent of students in NC 306 answered the question correctly. NC 306 therefore

underperformed on the seemingly easier item, but fared significantly better than similar unsuspecting schools on the more difficult items. Further cross-question variation can be seen over items 6, 7 and 8. The suspicious blocks of item responses and steep variation across items indicate that some form of cheating or suspicious behaviour may have taken place during the assessment.

This next school displays the worst M3 value which indicates a high cross question variance. This KwaZulu-Natal school has test data on 12 learners and appears to have done relatively well with students on average achieving a score of 17.5 out of 27. As is evident through the graph, this school shows extremes in the percentages of students with items correct. Either almost all students got the item in question correct or almost all students got the item incorrect, which indicates cheating may have taken place. The similar schools' average is based on 13 schools which achieved an average score of between 17 and 18.

Figure 8: Illustration of Strange School KZN 307



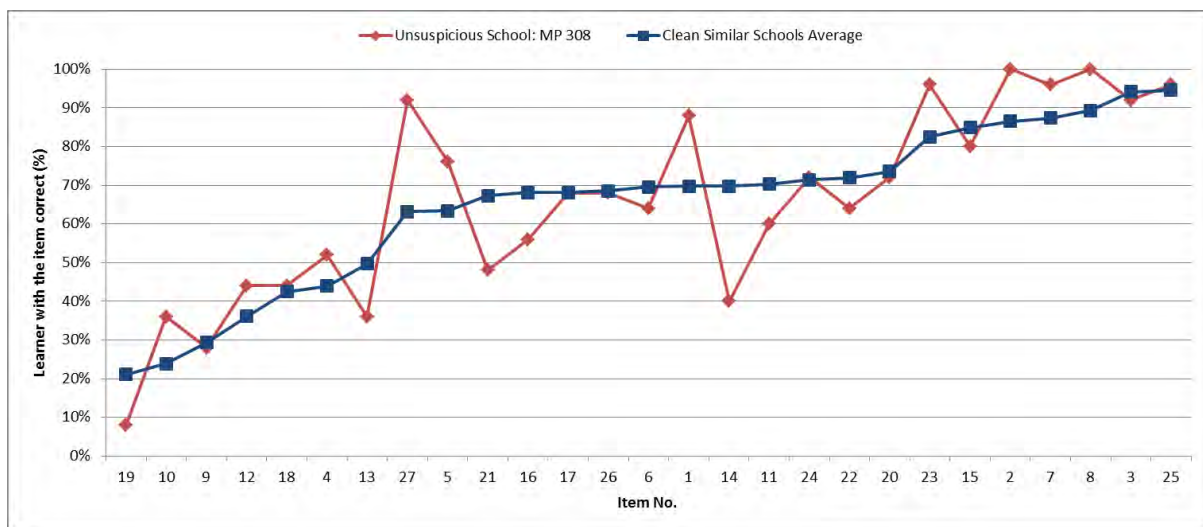
Source: V-ANA (2013)

There are a number of item responses in this school which reflect a high cross-question variance. At a glance a suspiciously high number of items reflect extreme responses in this school. All 12 learners in this school got 7 items correct when on average similar clean schools may have had 1 or 2 items where all students answered correctly. Similarly, there were 5 items where almost all students answered incorrectly in KZN 307 compared to an average of 1 item in similar schools. Specific examples showing the high degree of cross-question variance can be seen over a number of items. Items 11, 12 and 13 refer to a three-part question covering easy, moderate and difficult questions. Over these items the student response pattern swings from beating the similar clean

schools average with all students answering item 11 correctly, to item 12 where not a single student gets the answer right to item 13 where KZN 307 once again fairs better than similar unsuspecting schools. Further examples to this effect involve items 15 and 16 as well as items 22 and 23. This school shows a dubious number of extreme responses resulting in an overall divergence from similar unsuspecting schools of 30 percent.

Figure 9 illustrates the response pattern of a school which reports no suspicious activity according to the cheating measures developed. This graph compares the average student responses from 25 students in MP 308 who on average received a score of 17.76 out of 27 with the average student responses of 13 similar clean schools who achieved a score of between 17 and 18. This school diverges from the similar schools' average to a much lower degree than those schools displaying high suspicious measures. In addition this clean school follows the general pattern observed for similar unsuspecting schools which is expected given the very low cheating measure values observed.

Figure 9: Illustration of Non-Strange School MP 308



Source: V-ANA (2013)

With regards to the Language assessment, many of the suspicious schools outperformed the average over the same items such as items 9 and 10 and items 12 and 13. However, the suspicious schools varied over the items on which they underperformed. Overall suspicious schools seem to have fared worse on comprehension based items such as item 5, 6 and 27. Some of the items over which suspicious schools deviated most from the similar schools' average are displayed below.

Box 3: Grade 3 Language Items on which Suspicious Schools Outperformed the Average

Item 9 & 10: Question 8.1 & 8.2

- 8) Rewrite the sentences in the given tenses.
- 8.1) Past tense.
The farmer feeds the rabbits, while the children watch.
- 8.2) Present tense.
The pig played in the mud, while the horses galloped in the field.

Item 11, 12 & 13: Question 9.1 & 9.2 & 9.3

9. Read:
- 9.1) A noun is a naming word. Circle two nouns in the sentence below.
The dog wags its tail.
- 9.2) A pronoun is a word in the place of a noun. Circle two pronouns in the sentence below.
The farmer and his wife love the animals on their farm.
- 9.3) A conjunction is a joining word. Circle one conjunction in each of the sentences below.
- a. The pigs ate their food and went to sleep.
- b. The horses neigh, but the ducks quack.

Source: ANA (2013), Department of Basic Education

Box 4: Grade 3 Language Items over which Suspicious Schools Underperformed

Item 5 & 6: Question 5 & 6 (comprehension based)

- 5) "Nkosi Sikelel iAfrika" is our National Anthem. What made Enoch write this song?
Enoch wrote "Nkosi Sikelel iAfrika" because ...
- 6) In your opinion, why did Enoch become famous?
Enoch became famous because ...

Item 27: Question 15

- 15) Write two paragraphs of at least 10 sentences in total about a visit to a farm. Make sure that you use the correct punctuation, grammar and spelling. Do not number your sentences.

Source: ANA (2013), Department of Basic Education

5.2.4 Summary

The cheating measures here discussed were created and adapted from the Jacob and Levitt method in order to suit the ANA data. As such the previous selection of graphs were used to reveal whether these measures reflect in the school response patterns that which they are meant to reflect given their cheating indicator values. These graphs suggest that the suspicious indicators developed do allow for a meaningful interpretation of the data relating to the extent of cheating in national assessments.

Previous analysis of the V-ANA data by Gustafsson (2014) suggests that in the South African context, M2 and M3 measures may reflect the cleanliness of the data more than indicating potential cheating. However, the response patterns from graphs with high values for these measures clearly show that this is not the case. Those schools with high M2 values reflect suspicious blocks of answer strings whilst those schools with high M3 measures do display a high cross-question variance. These observations are particularly clear when there are no other confounding suspicious measure values.

Higher M4 values and school response patterns indicate that suspicious behaviour is more prevalent in Mathematics assessments compared to Language. M4 values, which show an overall degree of the implausibility of a schools average item response pattern, are almost halved for the Home Language assessment compared to the values seen for Mathematics. This results in 12 schools being identified as extremely suspicious for Mathematics whilst only 8 meet these criteria for Home Language in Grade 3. Overall it is evident that through these adapted measures, it is now possible to detect suspicious or outright cheating behaviour through a school's average student response pattern.

5.3. Detailed Analysis of Grade 6

This section investigates and discusses the prevalence of cheating in Grade 6 Mathematics and Language assessments with respect to the V-ANA and SACMEQ data. SACMEQ is the southern and eastern Africa consortium for monitoring educational quality. This study covers the conditions of schooling and quality of education in a host of southern and eastern African countries using reading, math and health assessments (Moloi & Chetty, 2011). South Africa participated in the SACMEQ project for the second time in 2007. SACMEQ assessments are carried out only for Grade 6 students and are designed according to the curriculum which allows for a direct comparison with the results of the Grade 6 South African ANA. One would expect

the SACMEQ data to display very low levels or no evidence of cheating due to the constant methodology and instruments applied across years (Moloi & Chetty, 2011). Furthermore, the SACMEQ tests are marked centrally in each province rather than in each school or district as has been the case for the ANAs.

Comparison between these two sets of assessments is informative as it gives an indication of what the ANA data ought to look like. It is important to note that some of the trends displayed in the SACMEQ data may differ from those shown in the ANA data. This is most likely due to the six-year gap between the 2007 SACMEQ data and the 2013 ANA data. Both the SACMEQ and ANA Mathematics and Language assessments are administered in only English and Afrikaans. The SACMEQ assessments, however, include only multiple choice questions whilst the ANAs vary in difficulty ranging from simple one mark answers or multiple choice questions to more difficult comprehensive or computational questions. Results for both Mathematics and Language are reported and compared.

5.3.1 Summary Statistics for V-ANA and SACMEQ Cheating Measures

The following table shows the summary statistics of cheating measures for Grade 6 Mathematics. A total of 387 and 392 schools are discussed for Mathematics with regards to the V-ANA and SACMEQ results respectively. Most of the cheating indicators reported below are lower for SACMEQ than for ANA which suggests either a greater degree of cheating in the ANA or a difference in the methodology and instruments used between the SACMEQ and ANA assessments.

Table 11: Summary Statistics on Cheating Measures for Grade 6 Mathematics

Assessment	Mean	SD	Min	10th %	Median	90th %	Max	Noted Schools
Mathematics Grade 6								
M2 ANA	0.35	0,21	0.01	0.16	0.31	0.53	2.06	53
M3 ANA	1.00	1,74	0.00	0.09	0.57	1.94	20.51	55
M4 ANA	-0.29	15,93	-29.61	-15.49	-1.92	15.27	76.28	39
M2 SACMEQ	0.22	0,09	0.08	0.15	0.21	0.30	1.36	42
M3 SACMEQ	0.34	0,71	0.03	0.10	0.20	0.55	12.79	50
M4 SACMEQ	-1.13	10,96	-46.25	-13.40	-1.83	11.84	60.47	49

Source: V-ANA (2013) & SACMEQ (2007)

The M2 and M3 values reported above are lower for SACMEQ data than V-ANA data. In particular, SACMEQ data seems to have a narrower band of values associated with a lower

standard deviation. The divergence in values is greater for M3 values than for M2 values with this divergence increasing towards the top end of the distribution. The 90th percentile values reflect this starkly as the M3 90th percentile value is 1.94 for the V-ANA data compared to a value of 0.55 for SACMEQ data. In fact the 90th percentile M2 and M3 values for SACMEQ are lower than the median M2 and M3 values for the V-ANA data. With regards to M4 values, the SACMEQ values lie significantly and consistently below those for the V-ANA data. These results show that SACMEQ data presents fewer suspicious blocks of responses as well as a lower overall degree of correlation in student answers across the test and a lower *variance* in the degree of correlation *across* questions compared to the Annual National Assessments. Looking specifically at the M4 values, approximately 10 percent of schools show suspicious behaviour in the V-ANA data.

Table 12 shows the summary statistics of cheating measures for Grade 6 Home Language. A total of 134 and 391 schools are discussed for this subject with regards to the V-ANA and SACMEQ results respectively. Due to the significantly smaller sample represented in the V-ANA data, Grade 6 Language values for SACMEQ data will be discussed with regards to both ANA Home Language (134 schools) and ANA First Additional Language (267 schools) in order to gain a more rounded view of the prevalence of cheating within the V-ANA data. Almost all summary values in the Language assessment are lower for the SACMEQ data compared to the V-ANA data which corroborates the findings of the summary Mathematics analysis.

Table 12: Summary Statistics on Cheating Measures for Grade 6 Language

Assessment	Mean	SD	Min	10th %	Median	90th %	Max	Schools
Home Language								
M2 ANA	0.24	0.11	0.00	0.11	0.23	0.38	0.61	14
M3 ANA	0.55	0.65	0.00	0.06	0.32	1.36	4.59	13
M4 ANA	-0.40	11.69	-49.61	-12.01	0.00	12.80	35.41	14
First Additional Language								
M2 ANA	0.41	0.14	0.12	0.25	0.39	0.58	1.05	26
M3 ANA	1.12	1.01	0.10	0.29	0.79	2.36	5.33	29
M4 ANA	1.21	12.92	-28.47	-14.06	-0.33	18.37	52.76	21
Language								
M2 SACMEQ	0.21	0.07	0.08	0.14	0.20	0.29	0.88	51
M3 SACMEQ	0.29	0.45	0.02	0.09	0.21	0.48	9.17	53
M4 SACMEQ	-0.74	8.76	-27.09	-11.69	-0.77	9.26	40.43	46

Source: V-ANA (2013) & SACMEQ (2007)

The suspicious indicator values presented above differ substantially between Home Language and First Additional Language within the V-ANA data. The V-ANA Home Language assessment is unevenly weighted with some provinces such as Limpopo only reflecting one school whereas other provinces such as the Western Cape reflect significantly more schools, 34. This would lower the suspicious measure values of the V-ANA data as provinces shown to reflect suspicious behaviour report a lower school sample whilst typically unsuspecting provinces reflect large school samples. The First Additional Language V-ANA results therefore provide a more accurate reflection of suspicious behaviour in the Language assessment due to the greater sample size and the more even distribution of schools. This subject will henceforth be used for comparison against the SACMEQ Language assessment.

SACMEQ values are on average lower than those displayed for the V-ANA data. M2 SACMEQ values are almost half that of the corresponding V-ANA values whilst M3 SACMEQ values are approximately a quarter of those seen in the corresponding V-ANA data. In addition, SACMEQ values reflect a more gentle gradient and on average only increase sharply past the 95th percentile. The SACMEQ data therefore shows lower levels of suspicious behaviour across all cheating indicators. Looking specifically at M4 values, there are 21 and 46 schools which exceed the national 90th percentile for the V-ANA and SACMEQ data respectively. Based on this criteria the SACMEQ data does reflect a greater number of suspicious schools, however the national 90th percentile values on which these values are based are half of those seen for the V-ANA data. The following tables, tables 13 and 14, show combinations of high measure values for Grade 6 Mathematics and Language subjects. Schools exceeding the 90th percentile values for the individual cheating measures, or composite measure, were considered to have high measure values and were assigned a value of “1”. These tables provide a cursory check and display a more conservative estimate of the prevalence of cheating in the ANA and SACMEQ data.

If one assumes that only those schools exhibiting high measure values across all cheating measures are engaged in suspicious behaviour then only 4.13 percent and 2.3 percent of schools can be labelled as “cheater schools” in the V-ANA and the SACMEQ data for Mathematics respectively. This is most likely the lower bound of the prevalence of cheating in Grade 6 Mathematics. The extent of suspicious behaviour in the V-ANA data is almost double that experienced in the SACMEQ data. Assessing the combinations of high measure values for the Language assessment, the V-ANA First Additional Language data presents the highest prevalence of suspicious behaviour, 2.62 percent compared to 0.77 percent found in the

SACMEQ data. As seems common throughout the grades, suspicious behaviour is more prominent in Mathematics than Language.

Table 13: Combinations of High Value Cheating Measures for Grade 6 Mathematics

Suspicious Answer String Measures				Noted Schools			
M2	M3	M4	JL3	ANA	(%)	SACMEQ	(%)
0	0	0	0	294	75.97	282	71.94
0	0	1	0	8	2.07	19	4.85
1	1	0	1	0	0	3	0.77
1	1	1	1	16	4.13	9	2.3
0	1	0	0	14	3.62	18	4.59
1	0	0	0	6	1.55	6	1.53
1	1	0	0	25	6.46	17	4.34
0	0	1	1	9	2.33	16	4.08
Other less frequent combinations				15	3.88	22	5.61
Total Schools				387		392	

Source: V-ANA (2013) & SACMEQ (2007)

Table 14: Combinations of High Value Cheating Measures for Grade 6 Language

Suspicious Answer String Measures				Noted Schools			
M2	M3	M4	JL3	ANA (FAL)	(%)	SACMEQ	(%)
0	0	0	0	210	78.65	301	76.98
0	0	1	0	4	1.5	19	4.86
1	1	0	1	0	0	1	0.26
1	1	1	1	7	2.62	3	0.76
0	1	0	0	10	3.75	14	3.57
1	0	0	0	7	2.62	5	1.28
1	1	0	0	9	3.37	19	4.85
0	0	1	1	5	1.87	8	2.04
Other less frequent combinations				15	5.62	21	5.36
Total Schools				267		391	

Source: V-ANA (2013) & SACMEQ (2007)

5.3.2 Analysis of Cheating Measures by Province

At a summary glance, the national perspective previously discussed shows that the prevalence of cheating is on average the same for Grade 3 and Grade 6; however, the cheating measure values reported are higher for Grade 6 in both Mathematics and Language. This national average however, may be hiding provincial trends as was the case with the Grade 3 results. This section will analyse whether all provinces are affected by cheating or suspicious behaviour to the same extent.

The following series of tables reflect the extent of suspicious behaviour by province for Mathematics and Language with respect to the V-ANA and SACMEQ data. Tables 15 to 19 display the prevalence of suspicious behaviour by province for the Mathematics assessment whilst tables 20 to 24 display the prevalence of suspicious behaviour by province for the Language assessment. Overall the SACMEQ data displays lower average suspicious measure values compared to the V-ANA data; however, the V-ANA and SAQMEQ data display different trends. This may indicate that there is either a significant difference between the targeting of these two assessments or that the distributions within provinces have changed drastically since 2007. Suspicious measure values report stark contrasts between provinces and once again these values are generally higher for Mathematics than for Language.

Table 9a shows the actual 90th percentile values of the cheating measures by province for the V-ANA and SACMEQ data with respect to the Mathematics assessment.

Table 15: 90th Percentile Cheating Measure Values by Province for Grade 6 Mathematics

Province	ANA				SACMEQ			
	M2	M3	M4	N	M2	M3	M4	N
EC	0,63	4,59	13,93	39	0,52	1,85	21,13	50
FS	0,59	2,39	19,9	40	0,31	0,54	9,3	40
GP	0,54	1,08	46,76	51	0,23	0,32	5,85	43
KZN	1,07	4,98	26,54	42	0,27	0,5	10,93	64
LP	0,56	2,03	4,94	42	0,27	0,4	18,27	40
MP	0,41	1,04	7,34	42	0,27	0,39	16,07	37
NC	0,46	1,33	4,3	44	0,29	0,63	11,85	40
NW	0,61	2,16	11,83	43	0,3	0,81	4,45	38
WC	0,43	1,81	6,5	44	0,33	0,74	9,68	40
National 90th Percentile	0,53	1,94	15,27	387	0,3	0,55	11,84	392

Source: V-ANA (2013) & SACMEQ (2007)

Within Mathematics the V-ANA and SAQMEQ data report almost opposite trends. The V-ANA data reflects highly suspicious behaviour for Gauteng, KwaZulu-Natal and the Free State. M4 values, for these provinces are particularly high, in some cases three times the national 90th percentile values as is the case for Gauteng. KwaZulu-Natal and the Free State reflect values above the national 90th percentile across all three cheating measures indicating exceptionally suspicious item response patterns in these provinces. The SACMEQ data, however, shows high M4 values for the Eastern Cape, Limpopo and Mpumalanga with the Eastern Cape being the only province to exceed the national 90th percentile across all three measures. The national 90th percentile values are much lower in the SACMEQ data suggesting a much lower level of suspicious behaviour on these assessments.

The 90th percentile values reported above have been used to present tables 9b to 9e, which show these results in a more accessible manner. Table 16 and 17 present the number and percent of schools which achieved cheating measure values above the national 90th percentiles whilst tables 18 and 19 report the number and percent of these noted schools with cheating measure values at least two standard deviations above the national 90th percentile values, by province.

Addressing the V-ANA results; 6 provinces display between 2 and 4 noted or suspicious schools which translates to between 4 percent and 10 percent of schools respectively. Whilst there seems to be wide spread suspicious behaviour in schools, only 1 school in Mpumalanga has cheating measure values greater than two standard deviations above the national 90th percentiles across all measures. There is significant variation between the cheating measures indicating that provinces may systematically use the same method to alter student assessment scores. Gauteng reflects 13 noted schools, 25 percent, with implausible school-level item averages compared to academically similar schools (measure 4) compared to only 4 noted schools, 8 percent, with suspiciously high within class variation (measure 3). Similarly KwaZulu-Natal shows 12 noted schools, 28 percent, with high M3 values compared to 5 noted schools, 12 percent, with exceptionally high M4 values.

Addressing the SACMEQ results; 5 provinces display noted schools, however, only the Eastern Cape shows excessive suspicious behaviour with 7 schools, 14 percent, displaying values greater than the national 90th percentile. Strangely the SACMEQ results show high measure values for the Western and Northern Cape as well as Gauteng and the Free State, most likely showing the time delay between the SACMEQ III assessment in 2007 and the ANA in 2013. Tables 18 and 19 further confirm that the Eastern Cape shows the most suspicious data with 16 percent, 12

percent and 2 percent of schools having cheating measures at least two standard deviations above the national 90th percentile for M2, M3 and M4 respectively.

Table 16: Number of Noted Schools by Province for Grade 6 Mathematics

Province	ANA					SACMEQ					
	M2	M3	M4	All	N	M2	M3	M4	All	N	
EC	9	9	4	3	39	14	18	10	7	50	
FS	5	6	7	3	40	4	3	2	1	40	
GP	7	4	13	2	51	1	1	3	1	43	
KZN	11	12	5	4	42	6	7	6	0	64	
LP	7	7	3	2	42	1	0	10	0	40	
MP	3	4	3	2	42	2	1	8	0	37	
NC	4	3	2	0	44	3	7	5	1	40	
NW	4	6	1	0	43	3	6	2	0	38	
WC	3	4	1	0	44	7	7	3	1	40	
Total	53	55	39	16	387	41	50	49	11	392	
National 90th Percentile Cut-offs Applicable:											
M2					0.53	M2					0.30
M3					1.94	M3					0.55
M4					15.27	M4					11.84

Source: V-ANA (2013) & SACMEQ (2007)

Table 17: Percent of Noted Schools by Province for Grade 6 Mathematics

Province	ANA					SACMEQ					
	M2	M3	M4	All	N	M2	M3	M4	All	N	
EC	23	23	10	8	39	28	36	20	14	50	
FS	13	15	18	8	40	10	8	5	3	40	
GP	14	8	25	4	51	2	2	7	2	43	
KZN	26	29	12	10	42	9	11	9	0	64	
LP	17	17	7	5	42	3	0	25	0	40	
MP	7	10	7	5	42	5	3	22	0	37	
NC	9	7	5	0	44	8	18	13	3	40	
NW	9	14	2	0	43	8	16	5	0	38	
WC	7	9	2	0	44	18	18	8	3	40	
National 90th Percentile Cut-offs Applicable:											
M2					0.53	M2					0.30
M3					1.94	M3					0.55
M4					15.27	M4					11.84

Source: V-ANA (2013) & SACMEQ (2007)

Table 18: Number of Schools at least 2 Standard Deviations above the National 90th Percentile for Grade 6 Mathematics

Province	ANA					SACMEQ					
	M2	M3	M4	All	N	M2	M3	M4	All	N	
EC	5	4	0	0	39	8	6	1	1	50	
FS	0	0	1	0	40	1	3	1	1	40	
GP	0	0	6	0	51	0	0	0	0	43	
KZN	7	6	0	0	42	0	0	1	0	64	
LP	2	1	0	0	42	0	0	0	0	40	
MP	1	2	1	1	42	0	0	1	0	37	
NC	1	1	1	0	44	0	0	0	0	40	
NW	0	1	0	0	43	0	1	0	0	38	
WC	0	0	0	0	44	0	0	0	0	40	
Total	16	15	9	16	387	9	10	4	2	392	
National Cut-offs Applicable:											
M2					0.94	M2					0.49
M3					5.43	M3					1.97
M4					47.13	M4					33.76

Source: V-ANA (2013) & SACMEQ (2007)

Table 19: Percent of Schools at least 2 Standard Deviations above the National 90th Percentile for Grade 6 Mathematics

Province	ANA					SACMEQ					
	M2	M3	M4	All	N	M2	M3	M4	All	N	
EC	13	10	0	0	39	16	12	2	2	50	
FS	0	0	3	0	40	3	8	3	3	40	
GP	0	0	12	0	51	0	0	0	0	43	
KN	17	14	0	0	42	0	0	2	0	64	
LP	5	2	0	0	42	0	0	0	0	40	
MP	2	5	2	2	42	0	0	3	0	37	
NC	2	2	2	0	44	0	0	0	0	40	
NW	0	2	0	0	43	0	3	0	0	38	
WC	0	0	0	0	44	0	0	0	0	40	
National Cut-offs Applicable:											
M2					0.94	M2					0.49
M3					5.43	M3					1.97
M4					47.13	M4					33.76

Source: V-ANA (2013) & SACMEQ (2007)

With regards to Language, table 20 as seen below, shows the actual 90th percentile values of the cheating measures by province for the V-ANA and SACMEQ data. The SACMEQ data seems to show a greater degree of suspicious behaviour compared to the V-ANA data. The SACMEQ findings, however, are based on much lower national 90th percentile values, values approximately half those observed for the V-ANA data. The V-ANA data shows highly suspicious behaviour for the Free State, Gauteng and KwaZulu-Natal. These provinces, with the exception of KwaZulu-Natal, exceed average 90th percentile values across all suspicious string indicators. The SACMEQ data, as with Mathematics, reports significantly lower suspicious measure values. The Eastern Cape, Limpopo and Western Cape display M4 values above the national 90th percentile, however only the Eastern Cape displays noted values across all measures.

Table 20: 90th Percentile Cheating Measure Values by Province for Grade 6 Language

Province	ANA First Additional Language				SACMEQ Language			
	M2	M3	M4	N	M2	M3	M4	N
EC	0,55	1,88	7,57	35	0,34	1,03	13,55	49
FS	0,69	3,96	31,6	35	0,27	0,41	8,79	40
GP	0,58	3,72	29,39	29	0,24	0,34	8,04	43
KZN	0,47	1,76	19,64	28	0,3	0,48	4,04	64
LP	0,55	1,77	2,31	40	0,28	0,43	19,68	40
MP	0,47	1,48	-0,61	35	0,29	0,67	8,68	37
NC	0,46	1,27	10,25	20	0,3	0,47	7,25	40
NW	0,57	2,29	15,24	35	0,27	0,4	7,93	38
WC	0,54	2	16,84	10	0,26	0,53	10,19	40
National 90 th Percentile	0.58	2.36	18.37	267	0.29	0.48	9.26	391

Source: V-ANA (2013) & SACMEQ (2007)

The 90th percentile values reported above have been used to present tables 21 to 24, which show these results in a more accessible manner. Table 21 and 22 presents the number and percent of schools which achieved cheating measure values above the national 90th percentiles whilst tables 23 and 24 report the number and percent of these noted schools with cheating measure values at least two standard deviations above the national 90th percentile values, by province. These tables corroborate the findings of table 20. Gauteng and the Free State are shown to be the most suspicious provinces with regard to the V-ANA data whilst the Eastern Cape displays the most suspicious cheating measure statistics in the SACMEQ data.

Strangely the V-ANA results display higher national 90th percentile values for Language compared to Mathematics. As such only 8 schools have cheating measure values greater than the

national 90th percentile values across all measures. Noted schools are evident in 5 provinces with the Free State and Gauteng displaying the worst results, 2 and 3 schools or 6 percent and 10 percent respectively. The Free State furthermore reports one school with cheating measure values greater than two standard deviations above the national 90th percentiles across all measures. The V-ANA data suggests that there was no systematic method of cheating on the language assessment but rather that if a province shows any form of suspicious behaviour then it mostly likely shows positive M2, M3 and M4 values.

The SACMEQ data follows the usual trend with lower suspicious values for the Language assessment compared to the Mathematics assessment. Overall 11 schools display suspicious data, 5 of which are in the Eastern Cape. The Eastern Cape is further confirmed as the most suspicious province with 3 schools, 6 percent, that report cheating measure values greater than two standard deviations above the national 90th percentiles across all measures. The SACMEQ data reveals the same trends for the Mathematics and Language assessments.

Table 21: Number of Noted Schools by Province for Grade 6 Language

Province	ANA					SACMEQ					
	M2	M3	M4	All	N	M2	M3	M4	All	N	
EC	4	2	1	0	35	16	16	9	5	49	
FS	6	6	5	2	35	3	1	3	0	40	
GP	4	7	6	3	29	0	1	5	0	43	
KZN	3	2	2	1	28	8	8	5	1	64	
LP	2	1	0	0	40	5	6	11	1	40	
MP	2	3	1	1	35	4	7	3	1	37	
NC	3	4	3	1	20	7	3	2	0	40	
NW	2	4	3	0	35	4	5	4	1	38	
WC	0	0	0	0	10	4	6	4	2	40	
Total	26	29	21	8	267	51	53	46	11	391	
National 90th Percentile Cut-offs Applicable:											
M2					0.58	M2					0.28
M3					2.36	M3					0.48
M4					18.37	M4					9.26

Source: V-ANA (2013) & SACMEQ (2007)

Table 22: Percent of Noted Schools by Province for Grade 6 Language

Province	ANA					SACMEQ					
	M2	M3	M4	All	N	M2	M3	M4	All	N	
EC	11	6	3	0	35	33	33	18	10	50	
FS	17	17	14	6	35	8	3	8	0	40	
GP	14	24	21	10	29	0	2	12	0	43	
KZN	11	7	7	4	28	13	13	8	2	64	
LP	5	3	0	0	40	13	15	28	3	40	
MP	6	9	3	3	35	11	19	8	3	37	
NC	15	20	15	5	20	18	8	5	0	40	
NW	6	11	9	0	35	11	13	11	3	38	
WC	0	0	0	0	10	10	15	10	5	40	
National 90th Percentile Cut-offs Applicable:											
M2					0.58	M2					0.28
M3					2.36	M3					0.48
M4					18.37	M4					9.26

Source: V-ANA (2013) & SACMEQ (2007)

Table 23: Number of Schools at least 2 Standard Deviations above the National 90th Percentile for Grade 6 Language

Province	ANA					SACMEQ					
	M2	M3	M4	All	N	M2	M3	M4	All	N	
EC	1	1	0	0	35	4	6	3	3	50	
FS	1	2	1	1	35	0	0	0	0	40	
GP	1	1	0	0	29	0	0	0	0	43	
KZN	1	0	0	0	28	2	1	0	0	64	
LP	0	0	0	0	40	0	2	2	0	40	
MP	1	0	1	0	35	0	2	0	0	37	
NC	0	2	0	0	20	0	0	0	0	40	
NW	0	1	0	0	35	0	1	0	0	38	
WC	0	0	0	0	10	0	0	0	0	40	
Total	5	7	2	1	267	6	12	5	3	392	
National Cut-offs Applicable:											
M2					0.86	M2					0.42
M3					4.38	M3					1.39
M4					44.20	M4					26.77

Source: V-ANA (2013) & SACMEQ (2007)

Table 24: Percent of Schools at least 2 Standard Deviations above the National 90th Percentile for Grade 6 Language

Province	Gr 6 ANA FAL					Gr 6 SACMEQ HL					
	M2	M3	M4	All	N	M2	M3	M4	All	N	
EC	3	4	0	0	35	8	12	6	6	50	
FS	3	6	3	3	35	0	0	0	0	40	
GP	3	3	0	0	29	0	0	0	0	43	
KZN	4	0	0	0	28	3	2	0	0	64	
LP	0	0	0	0	40	0	5	5	0	40	
MP	3	0	3	0	35	0	5	0	0	37	
NC	0	10	0	0	20	0	0	0	0	40	
NW	0	3	0	0	35	0	3	0	0	38	
WC	0	0	0	0	10	0	0	0	0	40	
National Cut-offs Applicable:											
M2					0.86	M2					0.42
M3					4.38	M3					1.39
M4					44.20	M4					26.77

Source: V-ANA (2013) & SACMEQ (2007)

The above tables display sharp contrasts between those provinces that do engage in cheating behaviour and those that don't. The V-ANA data displays higher national 90th percentile values for the Language assessment while the SACMEQ data follows the usual trend with higher measure values displayed for the Mathematics assessment. The V-ANA and SACMEQ data display different provincial trends in both the Mathematics and Language assessments most likely due to the 6 year time difference between the two assessments. Overall, the SACMEQ suspicious measure values presented were lower than those displayed for the V-ANA data.

With regards to the Mathematics assessment, KwaZulu-Natal, the Eastern Cape and the Free State are reported as the most suspicious provinces with respect to the V-ANA data. Over all suspicious measures, only the Eastern Cape exhibits highly suspicious results with regards to the SACMEQ data. The V-ANA data generally shows a broader spread of suspicious behaviour compared to the SACMEQ data which shows very concentrated cheating in the Eastern Cape. In the Language assessment, Gauteng and the Free State are shown to be the most suspicious provinces according to the V-ANA data whilst the Eastern Cape displays the most suspicious cheating measure statistics in the SACMEQ data.

For Mathematics overall only 4 percent of schools nationally show cheating measures above the 90th percentile across all measures compared to 2.8 percent for SACMEQ. For language 3

percent in both datasets have noted schools. This perspective hides stark provincial differences in V-ANA data with 10 percent of schools in KwaZulu-Natal and Gauteng reporting excessively high cheating values across all 3 measures in the Mathematics and Language assessments respectively. These values vary to greater degrees when considering individual cheating measures.

5.3.3 Patterns within Individual Schools

This section will explore individual suspicious schools and compare these to the average patterns displayed by similar clean or unsuspecting schools. In this instance similar clean schools refer to those unsuspecting schools which on average achieved the same overall score on the assessment as the suspicious school in question. Comparing the patterns of individual schools allows an assessment on whether these abstract adapted Jacob and Levitt measures are meaningful when applied to the V-ANA data, and how the V-ANA data differs to the patterns observed in the SACMEQ data.

The following series of graphs depict some of the most suspicious schools by cheating measure in comparison to similar clean schools. The items referred to on the horizontal axis are sorted according to the values of the similar schools' average. The perceived more difficult items are therefore displayed on the left of the graphs whilst the easier items are displayed to the right. Non-suspicious school patterns, those with low M2, M3, M4 and JL3 values, are expected to follow the similar schools' average quite closely whilst highly suspicious school patterns, those with high M2, M3, M4 and JL3 values, should diverge in varying degrees from the similar schools' average. The assigned suspicious school names give information on the province of the school, the grade, the order of the school being discussed and whether the school took part in the ANA or the SACMEQ assessment such as KZN 605 (S).

5.3.3.1 Mathematics

There are 16 schools in total that have excessively high suspicious measure values across all three cheating measures as well as the composite measure, JL3 in the Grade 6 V-ANA Mathematics data. The highest suspicious measure values for M2, M3, M4 and JL3 respectively are 2.06, 20.51, 67.76 and 4651.11 which are much higher than the values noted for Grade 3 Mathematics, in some cases double. The average assessment score associated with these suspicious schools is 22.9 compared to an average score of 16.08 for the 294 clean schools. Of the 16 schools that are in the top ten percentiles across all cheating measures, four are from KwaZulu-Natal, three each are

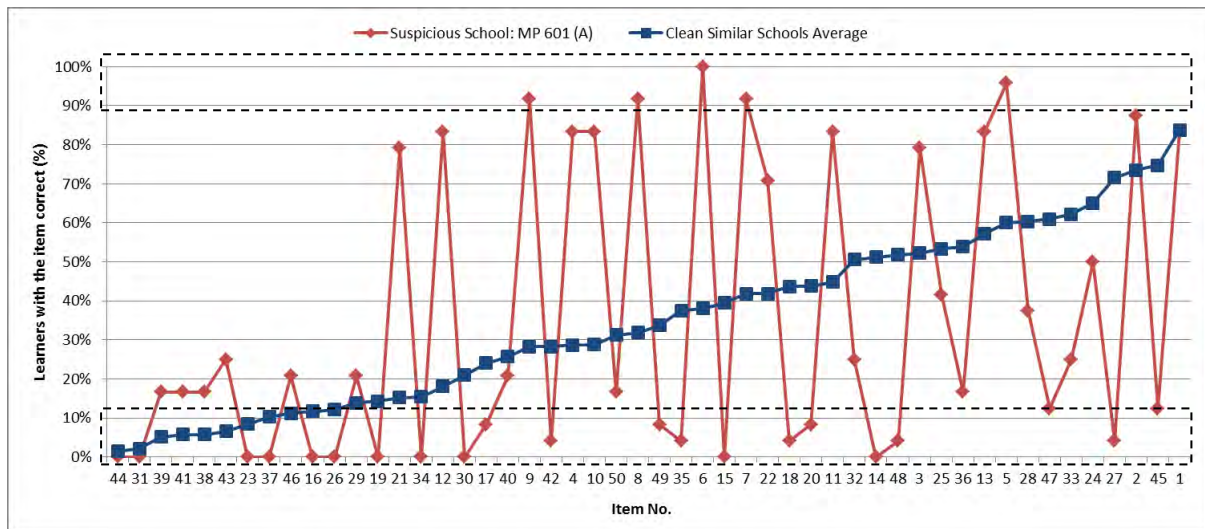
from the Eastern Cape and the Free State and two each are from Gauteng, Mpumalanga and Limpopo.

There are 9 schools with excessively high suspicious measure values across all cheating measures in the Grade 6 SACMEQ Mathematics data. The highest suspicious measure values for M2, M3, M4 and JL3 respectively are 1.36, 9.38, 60.47 and 3666.95 which are lower than the values noted for Grade 6 V-ANA Mathematics. The average assessment score associated with these suspicious schools is 20.67 compared to an average score of 18.08 for the 282 clean schools. Of the 9 schools that are in the top ten percentiles across all cheating measures, seven are from the Eastern Cape whilst one school each is from the Northern Cape and Free State.

Figures 10 and 11 below graph the most suspicious schools in the V-ANA and SACMEQ data according to M4 and JL3 values. Test data was available for 24 learners in the V-ANA suspicious school in Mpumalanga which on average achieved a score of 17 out of 50. The similar schools' average is based on 19 unsuspecting schools which achieved an overall average score of between 17 and 18 out of 50. Comparatively the suspicious SACMEQ school has data on 20 students who on average achieved a score of 25 out of 50. The SACMEQ similar schools average is based on 16 clean schools which on average achieved a score between 24 and 26 out of 50.

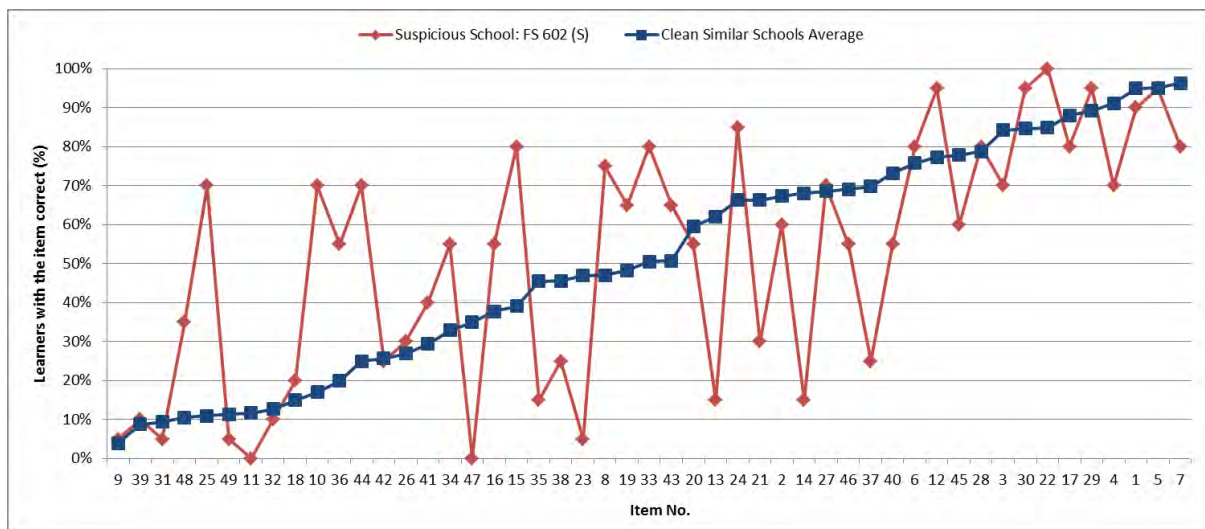
In both figures the suspicious school's response pattern diverges significantly from that of the similar schools' average as suggested by the high M4 values. This finding however is more prominent in the V-ANA suspicious schools with a much higher number of extreme responses compared to the SACMEQ suspicious school; there were 24 cases where less than 10 percent or greater than 90 percent of students in MP 601 (S) got the item correct. Over the assessment the V-ANA suspicious school on average deviates from similar school responses by 29 percent whilst the SACMEQ suspicious school on average deviates from similar school responses by 19 percent. The V-ANA suspicious school appears to display no trend, drastically outperforming or underperforming compared to the similar schools' average. The SACMEQ suspicious school tends to outperform similar schools on difficult questions and underperform on questions deemed easier according to similar schools.

Figure 10: Illustration of Strange School MP 601 (A)



Source: V-ANA (2013)

Figure 11: Illustration of Strange School FS 602 (S)



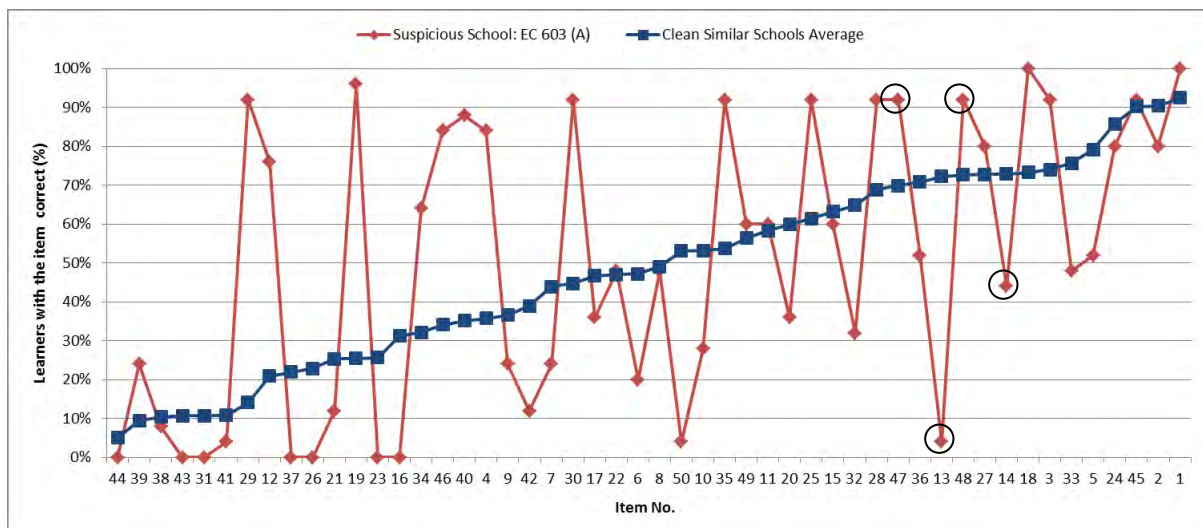
Source: SACMEQ (2007)

Figures 12 and 13 below graph the schools with very high M2 and M3 values with respect to the V-ANA and SACMEQ data. These schools should show suspicious blocks of item answers and a high cross-question variance. Test data was available for 25 learners in the V-ANA suspicious Eastern Cape school which on average achieved a score of 25 out of 50. The similar schools' average is based on 11 unsuspecting schools which achieved an overall average score of between 23 and 26. Comparatively the suspicious SACMEQ school has data on 17 students who on average achieved a score of 22 out of 50. The SACMEQ similar schools average is based on 16 clean schools which on average achieved a score between 20 and 22 out of 50. Both suspicious

blocks of response patterns and high cross-question variance are evident in the suspicious V-ANA and SACMEQ schools shown below. These suspicious responses are more pronounced in the V-ANA school who's M2 and M3 values are approximately double those of the SACMEQ school.

With respect to the V-ANA data, the suspicious Eastern Cape school seems to frequently beat the similar schools' average on some of the more difficult questions, towards the left of the graph, and underperform on questions deemed easier by students in similar schools, towards the right of the graph. Furthermore, the responses on items 13, 14, 47 and 48 showcase the dodgy nature of this schools data. The suspicious school performs significantly poorer than similar schools over items 13 and 14, but outperforms similar schools by approximately 20 percent over items 47 and 48. All these items are deemed to be relatively easy with over 70 percent of students in similar schools answering correctly. The large difference in correct responses over these items of relative equal difficulty is therefore unexpected and points to interference in the assessment or assessment scripts. There are also a number of instances where the percentage of learners with an item correct changes drastically over consecutive items and 22 cases of extreme responses, where less than 10 percent or greater than 90 percent of students answered an item correctly.

Figure 12: Illustration of Strange School EC 603 (A)

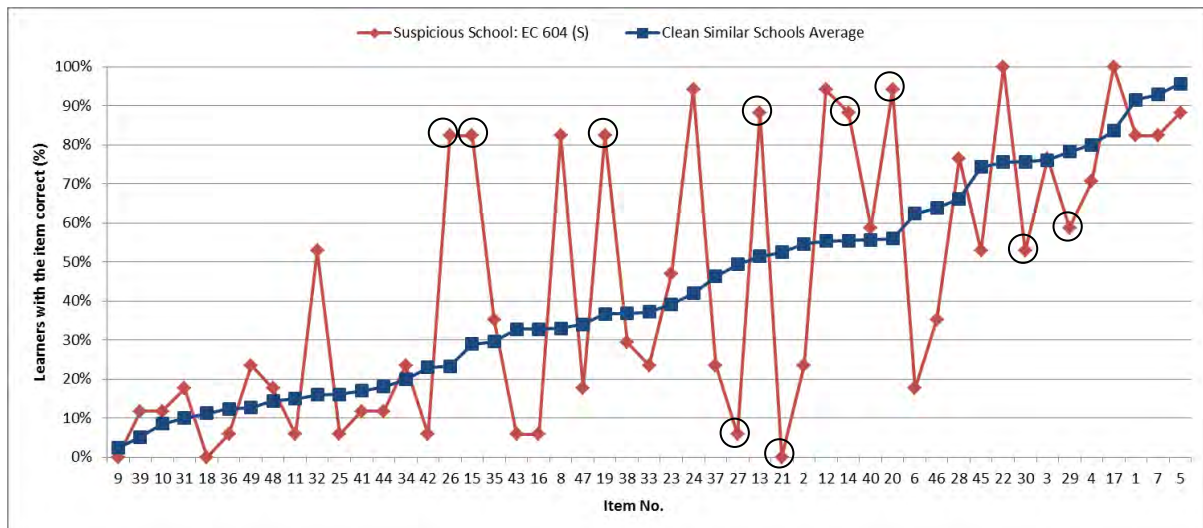


Source: V-ANA (2013)

With respect to the SACMEQ data, the suspicious Eastern Cape school seems to follow the trend of the similar schools' average at either end, on the hardest and easiest questions of the assessment, but diverges significantly over questions of average difficulty. The data displays a number of suspicious responses where the school in question either fared better or worse than

similar clean schools. Items 19 and 20 reflect the most suspicious block of responses as the suspicious school on average outperforms similar schools by approximately 40 percent to 50 percent. This is made more suspicious due to the varying degree of difficulty of these items. More than 80 percent of learners in the suspicious school answered these consecutive items correctly even though item 19 was deemed significantly more difficult than item 20 by similar performing students. Items 12, 13, 14 and 15 show another block of suspicious responses where students in EC 604 (S) significantly outperformed similar schools over questions of varying difficulty. Furthermore there are a number of instances where the percentage of learners with consecutive items correct changes drastically as seen over items 20 and 21 and 26 and 27. This school's response pattern follows the similar schools' average to a greater degree than that seen in the V-ANA data and reflects 15 cases of extreme responses compared to 22 respectively.

Figure 13: Illustration of Strange School EC 604 (S)



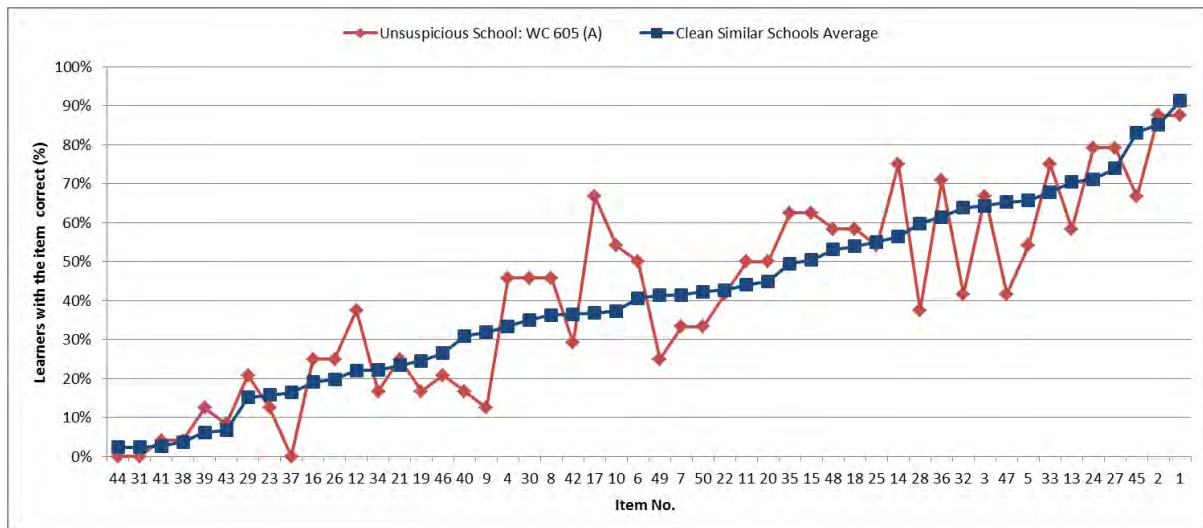
Source: SACMEQ (2007)

The following schools reflect unsuspecting responses, presenting very low values for the cheating measures created. The V-ANA Western Cape school achieved an average score of 20 out of 50 based on responses from 24 students. The similar schools' average is based on 12 unsuspecting schools which achieved an overall average score between 20 and 21 out of 50. The SACMEQ KwaZulu-Natal clean school achieved an overall average score of 24 out of 50 based on the responses of 23 learners. The similar schools' average is based on 17 clean schools which achieved an overall average score between 23 and 26.

These schools deviate only slightly from their respective clean similar schools average, 9 percent for the V-ANA school and 6 percent for the SACMEQ school. These schools follow the

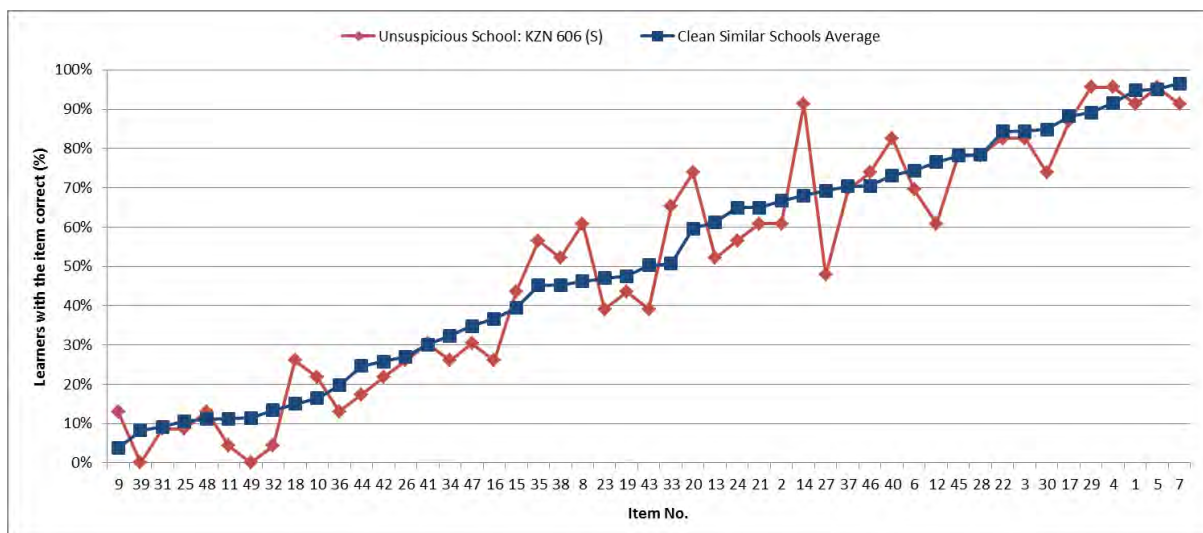
expected trend where almost all students are able to answer the easiest questions correctly, after which the percentage of correct student responses decreases as the questions become more difficult.

Figure 14: Illustration of Non-Strange School WC 605 (A)



Source: V-ANA (2013)

Figure 15: Illustration of Non-Strange School KZN 606 (S)



Source: SACMEQ (2007)

The following box shows some of the ANA items suspicious schools typically found easy or answered well. These items cover a range of skills including measurement, equations and number sequences or groups. Item 3 was deemed easy by similar clean schools whilst items 4 and 12 were seen as significantly more difficult. The same multiple choice SACMEQ assessments are used for each round and as such are not available to display in this paper.

Box 5: Grade 6 Mathematics Items on which Suspicious Schools Outperformed the Average

Item 3 and 4: Question 1.3 and 1.4

1.3) What is the length of the pencil shown below?



The length of the pencil is ...

- A) 56 cm
- B) 5,6 mm
- C) 0,56 m
- D) 5,6 cm

1.4) $22 + 35 = 63 - \Delta$, means that $\Delta = \dots$

- A) 6
- B) 28
- C) 41
- D) 57

Item 12: Question 3

3) Write down the next prime number.

19 , 23 , 29 , _____

Source: ANA (2013), Department of Basic Education

This next box displays some of the ANA items on which students of suspicious schools answered poorly. These items include numeric calculations, angles and price comparisons. Items 14 and 33 were deemed relatively easy whilst item 42 was perceived to be more difficult by similar clean schools. Over these items, schools with suspicious cheating values typically performed worse than similar unsuspecting schools.

Box 6: Grade 6 Mathematics Items over which Suspicious Schools Underperformed

Item 14: Questions 5.1

- 5.1) Calculate the answer:
 $43\,489 + 345\,987 + 307 =$

Item 33: Question 17.2

obtuse angle; acute angle; right angle

- 17) Use the words above to say what kind of angles are marked in the picture.



17.1 -----

17.2 -----

Item 42: Question 24

- 24) Examine the packets of sugar and their prices. Which is the best buy?



Source: ANA (2013), Department of Basic Education

5.3.4.2 Language

In total, 7 schools have excessively high suspicious measure values across all three cheating measures as well as the composite measure, JL3 in the Grade 6 V-ANA First Additional Language data. The highest suspicious measure values for M2, M3, M4 and JL3 respectively are

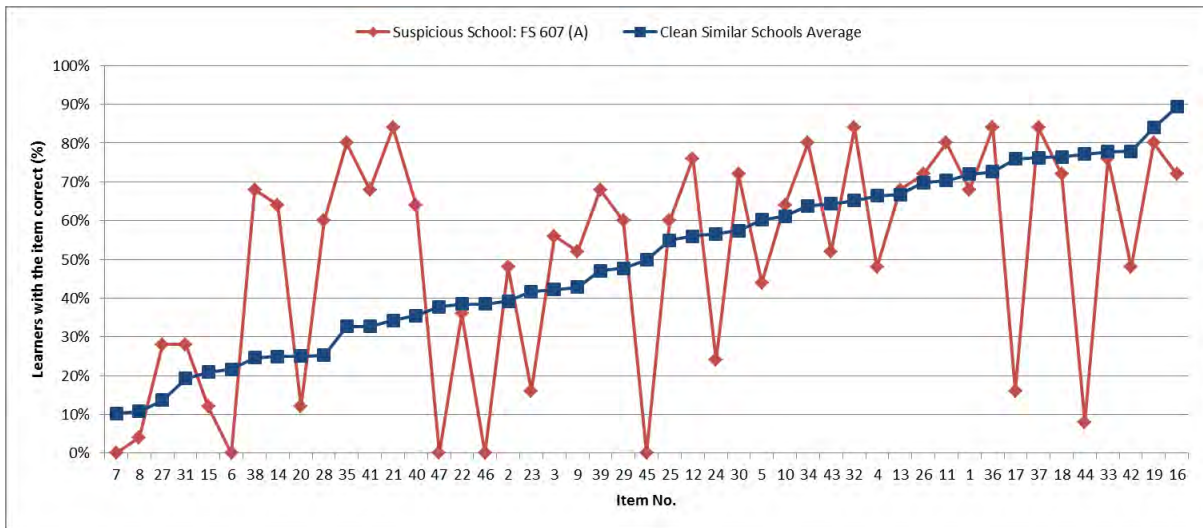
1.05, 4.53, 52.76 and 2804.67. These values are between 1.8 and 3.2 times the values seen in the V-ANA Grade 3 Home Language assessment based on M4 and JL3 measures indicating significantly more suspicious data in the Grade 6 results. The average assessment score associated with these suspicious schools is 25 compared to an average score of 20.41 for the 210 clean schools. Of the 7 schools noted across all cheating measures, three are from Gauteng with one each based in the Northern Cape, the Free State, Mpumalanga and KwaZulu-Natal.

There are 8 schools in total that have excessively high suspicious measure values across all three cheating measures as well as the composite measure, JL3 in the Grade 6 SACMEQ Language data. The highest suspicious measure values for M2, M3, M4 and JL3 respectively are 0.88, 9.14, 40.43 and 1653.54. The M2 and M3 values are very similar to those of the SACMEQ Mathematics assessment; however the M4 and JL3 values are almost half the value of those seen in Mathematics. This suggests that there are significantly fewer suspicious patterns in the SACMEQ language assessment compared to similar clean schools. The average assessment score associated with these suspicious schools is 30.41 compared to an average score of 27.98 for the 270 clean schools. Of the 8 schools that are in the top ten percentiles across all cheating measures, four are from the Eastern Cape, two are from the Western Cape and one school each is based in KwaZulu-Natal and Mpumalanga.

Figures 16 and 17 below graph the most suspicious schools in the V-ANA and SACMEQ data according to M4 and JL3 values. Test data was available for 25 learners in the Free State V-ANA suspicious school which achieved an average score of 23.4 out of 47. The V-ANA similar schools average is therefore based on 18 clean schools which scored between 23 and 24 on the assessment. Comparatively the SACMEQ suspicious school achieved an average assessment score of 29 out of 55 based on data from 11 students. The SACMEQ similar schools average is based on 10 clean schools which scored between 28 and 30 on the assessment.

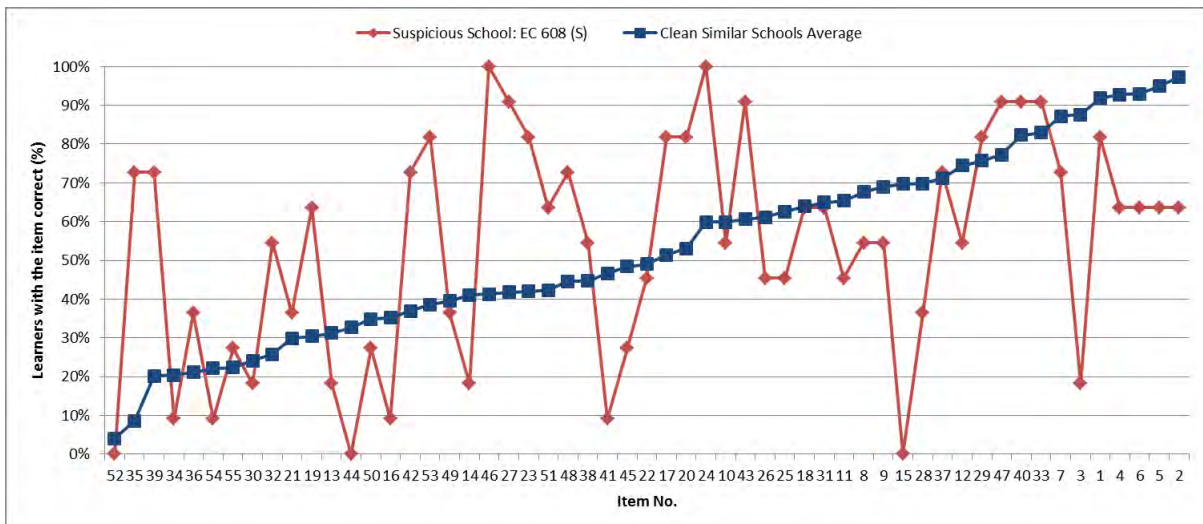
In both figures the suspicious school's response pattern diverges significantly from that of the similar schools' average. The V-ANA school however seems to display less suspicious responses, generally following the similar schools trend over questions considered by similar schools to be only moderately difficult. The more suspicious SACMEQ school deviates from the similar schools' average by approximately 23 percent and displays double the number of extreme responses, 14 as opposed to 7, observed in the V-ANA school. Overall both schools seem to outperform the average on difficult questions and underperform compared to similar schools on easier questions.

Figure 16: Illustration of Strange School FS 607 (A)



Source: V-ANA (2013)

Figure 17: Illustration of Strange School EC 608 (S)

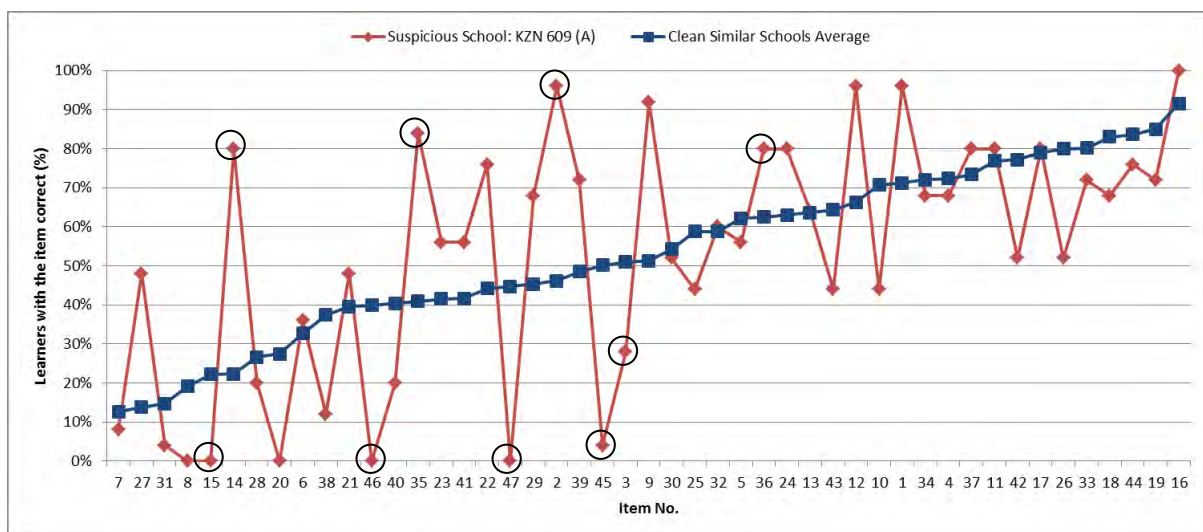


Source: SACMEQ (2007)

These next schools have very high M2 and M3 values with respect to the V-ANA and SACMEQ data. These schools should show suspicious blocks of item answers and a high cross-question variance. Test data was available for 25 learners in the V-ANA suspicious school from KwaZulu-Natal which on average achieved a score of 25 out of 50. The similar schools' average is based on 19 unsuspecting schools which achieved an overall average score of between 24 and 26. Comparatively the suspicious Eastern Cape SACMEQ school has data on 25 students who on average achieved a score of 47 out of 55. The SACMEQ similar schools average is based on 12 clean schools which on average achieved a score between 44 and 47 out of 50 as no other clean school scored so highly.

With respect to the V-ANA data, KZN 609 (A) appears to deviate more from the similar schools' average over the harder questions whilst broadly following the similar schools trend over easier questions. This school shows two stark examples of suspicious blocks of answer strings. Over items 45, 46 and 47 almost all students answer these consecutive items incorrectly whilst approximately 40 percent to 50 percent of students in similar unsuspecting schools answered correctly. Conversely the suspicious school fares significantly better than similar schools over items 35 and 36 with at least 80 percent of students answering correctly for both items despite the fact that item 35 was shown to be a much more difficult question. Furthermore there are a number of instances where the percentage of learners with an item correct changes drastically over consecutive items in some cases by as much as 68 percent to 80 percent, seen over items 2 and 3 and 14 and 15 respectively. The suspicious blocks of responses, high cross-question variance and extreme responses displayed in this school result in an overall deviation of 20 percent from the clean similar schools average.

Figure 18: Illustration of Strange School KZN 609 (A)

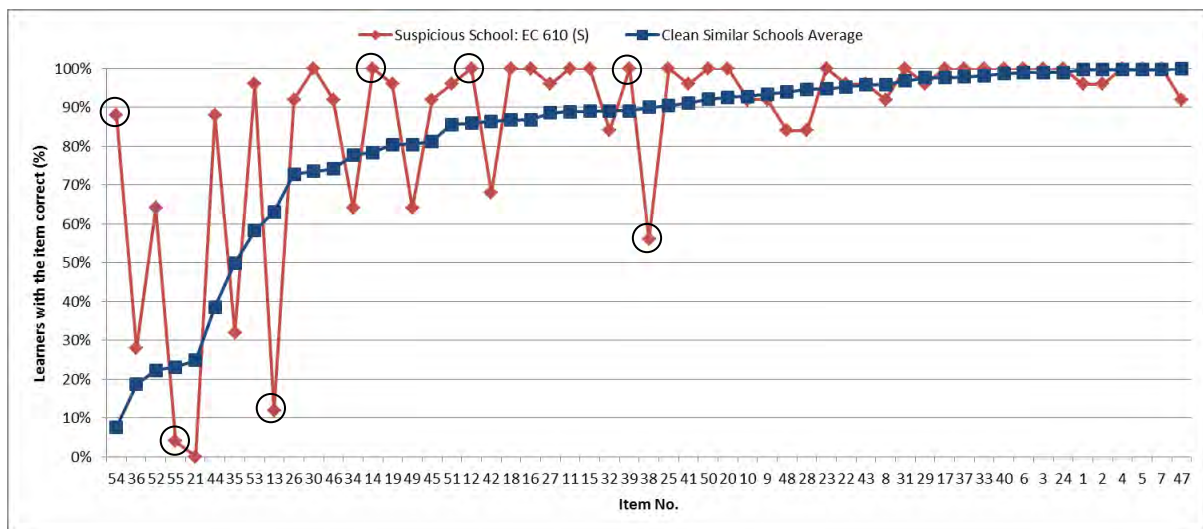


Source: V-ANA (2013)

The suspicious Eastern Cape SACMEQ school shown below reports a high M2 and M3 value despite the high concentration of overall correct responses. Due to the high academic performance of these schools, the suspicious school follows the similar schools trend within 10 percent over approximately half of the assessment. Given the high score achieved, this school reflects a high cross-question variance particularly over items 12, 13 and 14, 38 and 39 and 54 and 55. The high cross-question variance seen over items 38 and 39 and 54 and 55 of 44 percent and 84 percent respectively are odd given the similar perceived difficulty of these questions. Items 12, 13 and 14 display the yo-yo effect of a large variance jumping from 100 percent of

students answering item 12 correct to just 12 percent answering item 13 correct to 100 percent answering items 14 and 15 correctly. In addition the suspicious school has significantly outperformed the similar schools' average over some of the most difficult items. Despite the high academic performance displayed, EC 610 (S) deviates from the similar clean schools average by 13 percent.

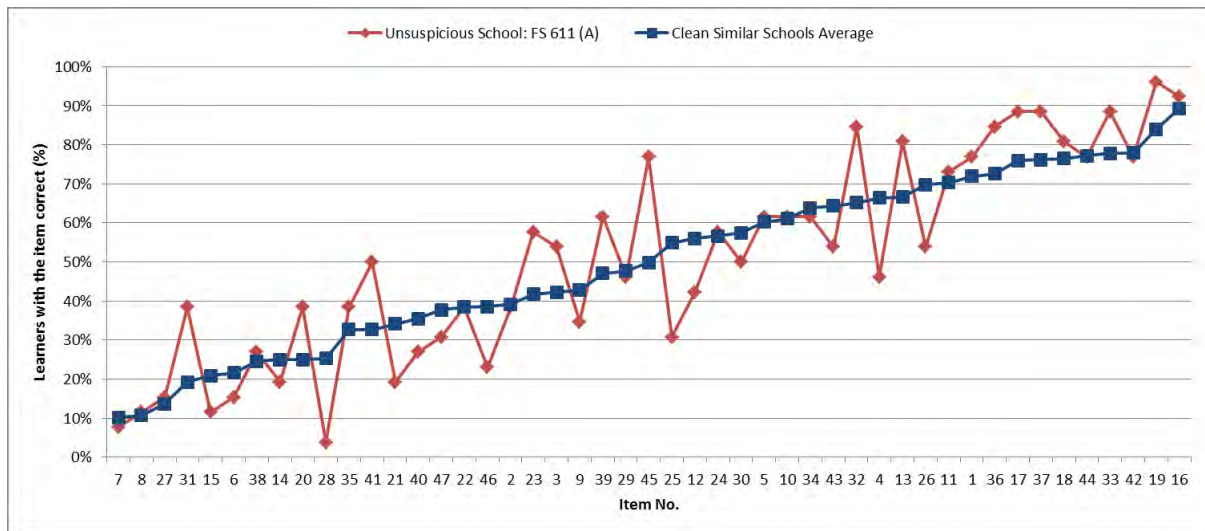
Figure 19: Illustration of Strange School EC 610 (S)



Source: SACMEQ (2007)

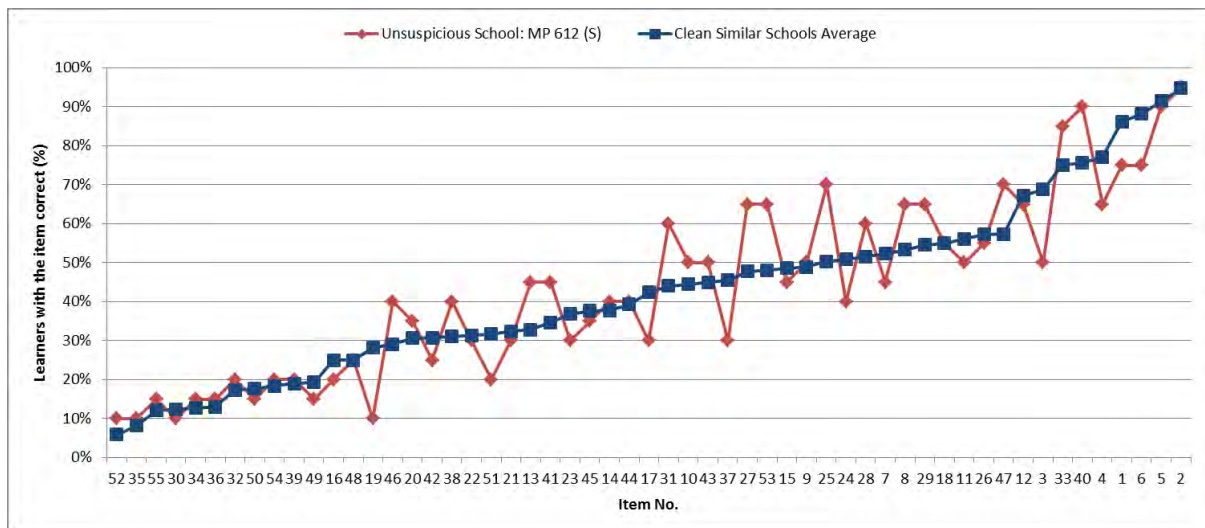
The following schools are shown to be clean, presenting very low values for the cheating measures created. The unsuspecting V-ANA school from the Free State achieved an overall average score of 24 out of 50 based on test data of 26 students. The similar clean schools average is based on 18 unsuspecting schools which achieved an overall score of between 23 and 24. The SACMEQ clean Mpumalanga school on average achieved an assessment score of 24 out of 55 based on 20 students. The similar schools' average is based on 28 unsuspecting schools which on average score between 23 and 24 out of 55. These schools follow the overall trend of similar schools with few students answering the difficult questions correctly and most students answering the easier questions correctly. Visually it is very easy to see that these schools do not display the erratic or systematic response patterns observed in the suspicious schools noted above. The response patterns of these clean schools deviate from the similar schools' average to a much smaller degree compared to suspicious schools, on average 9 percent in the V-ANA clean school and 7 percent in the SACMEQ clean school.

Figure 20: Illustration of Non-Strange School FS 611 (A)



Source: V-ANA (2013)

Figure 21: Illustration of Non-Strange School MP 612 (S)



Source: SACMEQ (2007)

Box 7 and 8 display some of the ANA items on which suspicious schools either beat the clean similar schools average or underperformed. The same multiple choice SACMEQ assessments are used for each round and as such are not available to display in this paper.

Box 7: Grade 6 Language Items on which Suspicious Schools Outperformed the Average

Item 14: Question 14 (Comprehension based)

- 14) Circle the letter of the correct answer.
What does 'the jackal flattered the crow' mean?
- A. He praised her more than was needed.
 - B. He gave her advice he thought would help.
 - C. He spoke to her softly and kindly.
 - D. He greeted her warmly.

Item 35: Question 30.2

- 30) Give the opposite gender of the underlined words.
30.2 The lion would not have done something like this to the jackal. _____

Item 39: Question 33

- 33) Rewrite the word in brackets by adding the apostrophe (').
The jackal stole the (crows) _____ cheese.

Item 41: Question 35

- 35) Identify the root/base word from the underlined word.
The queen of the birds looked very sad. _____

Source: ANA (2013), Department of Basic Education

Box 8: Grade 6 Language Items over which Suspicious Schools Underperformed

Item 15: Question 15 (Comprehension based)

- 15) Did the jackal like the crow's singing? Give a reason for your answer.

Item 31: Question 28

- 28) Write two sentences to show two different meanings of the word "back".

Items 45, 46 & 47: Question 38.1, 38.1 & 38.3

- 38) Write to your friend and tell him/her about the day you landed into trouble at school.
Write at least four well-linked paragraphs of 80 – 100 words in a letter format with the correct content. Make sure that you use the correct spelling, punctuation and grammar.
- Mark Allocation:
- 38.1) Format (5 marks)
 - 38.2) Language, Spelling and Punctuation (5 marks)
 - 38.3) Content (5 marks)

Source: ANA (2013), Department of Basic Education

5.3.4 Summary

As with the Grade 3 results, the Grade 6 cheating measures do reflect the suspicious response patterns they were designed to show. This suggests that suspicious behaviour can be investigated and discovered through assessment data and that these measures hold across grades, subjects and different assessment designs.

Higher suspicious measure values indicate that there may be significantly more interference in the ANA than in the SACMEQ assessments for both Mathematics and Language subjects. At a national level, the SACMEQ data reflects approximately 2 percent of schools as “cheater schools” across Mathematics and Language whilst the ANA reflect 2.6 percent to 4 percent of schools as “cheater schools” in the Language and Mathematics assessments respectively. These results vary greatly by cheating measure and province. In the V-ANA data 10 percent of Gauteng schools display highly suspicious response patterns in Mathematics whilst 10 percent of KwaZulu-Natal schools display highly suspicious response patterns in Language across all measures. Over all suspicious measures, only the Eastern Cape exhibits highly suspicious results with regards to the SACMEQ data.

M4 and JL3 values are higher for the Grade 6 data compared to the Grade 3 data which could indicate that cheating may become more robust the longer it goes undetected and as such is more evident in higher grades. Comparing the suspicious response patterns between Grade 3 and Grade 6 graphically, this seems to be the case. Further analysis involving higher grades, however, would need to be undertaken to assess this as it has been suggested by Gustafsson (2014) that high schools may reflect lower levels of suspicious data responses due to their experience of Grade 12 examinations.

5.4 Grade 9

The Grade 9 V-ANA results, whilst available for Mathematics and Language, pose a number of estimation challenges. Only 193 and 1078 students passed Grade 9 Mathematics and Home Language, much fewer students than seen in other grades. The low pass rates, especially for Mathematics, reduce the variation in student responses, amplify the effect of outliers and affect the accuracy of estimation. This is especially worrying when analysing the data at a provincial level as the distribution of pass rates is uneven across provinces. As such the Grade 9 results will not be discussed here.

5.5 Other Interpretations

Over Grades 3 and 6 those schools with highly suspicious response patterns on average achieve higher overall assessment scores when compared to clean schools. One would therefore expect that when a “cheater school” diverges from the similar schools’ average, they would outperform the similar schools’ average. There are however a number of instances where schools displaying suspicious data have underperformed drastically over a number of consecutive questions. School personnel and teachers would have no incentive to produce lower scores for students and would be expected to “cheat” only if the answers were known to be correct, thereby improving the assessment score. For those suspicious schools that underperform over dodgy responses, it may be that the individual or teacher cheating does not know the correct answer, as suggested in the DBE ANA report, or that perhaps this section of the curriculum hasn’t yet been taught.

Across Grade 3 and Grade 6 suspicious schools, especially those with high M2 values, the data was inspected for specific items where on average less than 30 percent of students answered the items correctly in Mathematics and Language. Looking at the response patterns, there are a number of items in each grade and subject which stand out as very poorly answered questions amongst the suspicious schools. Comparing these items to the ANA scripts however reveals that it is unlikely that poor performance is a result of untaught sections of the curriculum as these questions cover a range of skills and topics. This seems to confirm the original suspicion that these anomalies “may be driven by incorrect teachings within individual schools or classes” or that “test administration procedures may not have been properly carried out” (Department of Basic Education, 2011).

Chapter 6: Conclusion and Recommendations for Further Work

This analysis shows that the suspicious indicators developed do allow for a meaningful interpretation of the data relating to the extent of cheating in national assessments. Nationally, cheating seems to be prevalent in at least 10 percent of schools in Grade 3 and Grade 6. A Provincial breakdown reveals that suspicious school responses are noticeably higher in the Eastern Cape, the Free State and KwaZulu-Natal, and to some extent in Gauteng for Grade 6 assessments. The manner in which schools behave suspiciously varies significantly by province, subject and measure. The V-ANA data shows more pronounced cheating with higher cheating measure values than those seen in the SACMEQ assessment. Cheating seems to be more prominent in Mathematics assessments compared to Language as is noted in international literature on high stakes subjects.

This method is only possible on data that is recorded at the item-level, and a bigger effort should therefore be made to record at least a sub-sample of each educational dataset in order to improve access to such analysis. All suitably recorded educational datasets available such as TIMSS and PIRLS should be analysed for suspicious response patterns and not assumed to provide valid performance measures so as to improve upon or strengthen the findings presented here. This would help identify baseline cheating measure values, allow for a more detailed analysis on the determinants of cheating and add a further layer of validity to international or national assessment results. It would be particularly interesting in South Africa to apply these measures to a sub-sample of the Grade 12 school-leavers examination as these assessments involve high personal stakes for students and are regarded highly by potential universities.

With respect to the ANAs, it is vital that these assessments continue. The ANAs are an important tool in analysing the problems teachers and students face lower down in the educational system. By highlighting these problems in the early phases of education, corrective solutions can be implemented which will allow for effective “teaching and learning, improved literacy and numeracy at schools, better National Senior Certificate examination performance as well as expanding early childhood development” (Department of Basic Education, 2013, p.1). In future a concerted effort should be made to record the class data of these assessments in a uniform manner to ensure that specific classrooms and therefore teachers can be identified as cheater classrooms. It would also be useful if students could be tracked across years as this would allow the use of additional Jacob and Levitt cheating measures.

The literature suggests that high stakes testing and measurement distort intrinsic motivation and induce negative behavioural distortions. These unfortunate unintended consequences, however, can be seen even in no-stakes or low-stakes assessments like the Annual National Assessments. This suggests that the mere act of assessment and measurement is perceived as a stake by teachers and educational personnel. Better methods of assessment implementation need to be explored to improve testing and increase educational quality without compromising the autonomy and respect of educators.

Reference List

- Aghion, P., Caroli, E. and Garcia-Penalosa, C. (1999). "Inequality and Economic Growth: The Perspective of the New Growth Theories." *Journal of Economic Literature*, 37(4), pp. 1615 - 1660.
- Aiken, L. (1991). "Detecting, Understanding, and Controlling for Cheating on Tests." *Research in Higher Education*, 32(6), pp. 725 - 736.
- Akerlof, G.A. and Kranton, R.E. (2002). "Identity and Schooling: Some Lessons for the Economics of Education." *Journal of Economic Literature*, 40(4), pp. 1167 - 1201.
- Angoff, W.H. (1974). "The Development of Statistical Indices for Detecting Cheaters." *Journal of the American Statistical Association*, 69(345), pp. 44 - 49.
- Auditor General South Africa. (2015). "A Bold Vision for the Future: 2014 - 2015 Integrated Annual Report." Cape Town.
- Barro, R.. (1998). "Notes on Growth Accounting." *NBER Working Paper Series*, Volume 6654.
- Barro, R.J. and Lee, J.W. (2010). "A New Dataset of Educational Attainment in the World, 1950 - 2010." *National Bureau of Economic Research*, Volume No. W15902.
- Bellezza, F.S. and Bellezza, S.F. (1989). "Detection of Cheating on Multiple-Choice Tests by using Error-Similarity Analysis." *Teaching of Psychology*, 16(3), pp. 151 - 155.
- Bobba, M. and Coviello, D. (2007). "Weak Instruments and Weak Identification, in Estimating the Effects of Education, on Democracy." *Economics Letters*, Volume 96, pp. 301 - 306.
- Booher-Jennings, J. (2006). "Rationing Education in an Era of Accountability." *Phi Delta Kappan*, Volume 87, pp. 756 - 761.
- Borja, R.R. (1999). "Comments: SOLs raise concern, little support/change tests, use multiple criteria, speakers say." *Richmond Times-Dispatch*.
- Bruns, B., Filmer, D. and Patrinos, H.A. (2011). "Making Schools Work: New Evidence on Accountability Reforms." Washington, D.C.: The World Bank.
- Carnoy, M. and Loeb, S. (2002). "Does External Accountability Affect Student Outcomes? A Cross-State Analysis." *Educational Evaluation and Policy Analysis*, 24(4), pp. 305 - 331.
- Chiang, H. (2009). "How Accountability Pressure on Failing Schools Affects Student Achievement." *Journal of Public Economics*, Volume 93, pp. 1045 - 1057.
- Crittenden, V.L., Hanna, R.C. and Peterson, R.A. (2009). "The Cheating Culture: A Global Societal Phenomenon." *Business Horizons*, 52(4), pp. 337 - 346.
- Cullen, J.B. and Reback, R. (2006). "Tinkering Toward Accolades: School Gaming Under a Performance Accountability System." (No. w1228). National Bureau of Economic Research.
- Cutler, D.M. and Lleras-Muney, A. (2006). "Education and Health: Evaluating Theories and Evidence." (No. w12352). National Bureau of Economic Research.

- D'Addio, A.C. and d'Ercole, M.M. (2005). "Trends and Determinants of Fertility Rates: The Role of Policies." *OECD Social, Employment and Migration Working Papers*, Volume 27.
- Darling-Hammond, L. (1991). "The Implications of Testing Policy for Quality and Equality." *The Phi Delta Kappan*, 73(3), pp. 220 - 225.
- de Wolf, I.F. and Janssens, F.J. (2007). "Effects and Side Effects of Inspections and Accountability in Education: An Overview of Empirical Studies." *Oxford Review of Education*, 33(3), pp. 379 - 396.
- Deloitte and NEXIA SAB&T. (2014). "Report on Verification Annual National Assessment 2013 - Executive Summary."
- Department of Basic Education. (2011). "Report on the Annual National Assessments of 2011." Pretoria: Department of Basic Education.
- Department of Basic Education. (2013). "Report on the Annual National Assessment of 2013: Grades 1 to 6 & 9." Pretoria: Department of Basic Education.
- Elmore, R.F., Abelman, C.H. and Fuhrman, S.H. (1996). "The New Accountability in State Education Reform: From Process to Performance." In: H. Ladd, ed. *Holding Schools Accountable: Performance-based Reform in Education*. Washington, DC.: The Brookings Institution, pp. 65 - 89.
- Ferraz, C. and Bruns, B. (2011). "Incentives to Teach: The Effects of Performance Pay in Brazilian Schools." Washington, DC: World Bank.
- Ferraz, C. and Bruns, B. (2012). "Paying Teachers to Perform: The Impact of Bonus Pay in Pernambuco, Brazil." *Society for Research on Educational Effectiveness*.
- Figlio, D.N. and Lucas, M.E. (2000). "What's in a Grade? School Report Cards and House Prices." (No. w8019). *National Bureau of Economic Research*.
- Figlio, D.N. (2006). "Testing, Crime and Punishment." *Journal of Public Economics*, Volume 90, pp. 837 - 851.
- Figlio, D.N. and Getzler, L.S. (2002). "Accountability, Ability and Disability: Gaming the System." (No. w9307). *National Bureau of Economic Research*.
- Fuhrman, S. & Elmore, R.F. (2004). "Redesigning Accountability Systems for Education." *Critical Issues in Educational Leadership Series* ed. New York: Teachers College Press.
- Fullan, M. (2002). "The Three Stories of Education Reform." *Phi Delta Kappan*.
- Galloway, M.K. (2012). "Cheating in Advantaged High Schools: Prevalence, Justifications, and Possibilities for Change." *Ethics & Behaviour*, 22(5), pp. 378 - 399.
- Goertz, M.E. and Duffy, M.C. (2001). "Assessment and Accountability Systems in the 50 States: 1999 - 2000."
- Gustafsson, M. (2014). "A Check on Item-level Data Patterns in the 2013 ANA Associated with Possible Cheating."
- Haney, W. (2000). "The Myth of the Texas Miracle in Education." *Education Policy Analysis Archives*, 8(41), p. 41.

- Hanushek, E.A. (2002). "Publicly Provided Education." *Handbook of Public Economics*, Volume 4, pp. 2045 - 2141.
- Hanushek, E.A and Woessmann. (2008). "The Role of Cognitive Skills in Economic Development.." *Journal of Economic Literature*, 46(3), pp. 607 - 668.
- Hanushek, E.A. and Raymond, M.E. (2004). "The Effect of School Accountability Systems on the Level and Distribution of Student Achievement." *Journal of the European Economic Association*, 2(3), pp. 406 - 415.
- Heilig, J.V. and Darling-Hammond, L. (2008). "Accountability Texas-Style: The Progress and Learning of Urban Minority Students in a High-Stakes Testing context." *Educational Evaluation and Policy Analysis*, 30(2), pp. 75 - 110.
- Holland, P.W. (1996). "Assessing Unusual Agreement between the Incorrect Answers of Two Examinees using the K-Index: Statistical Theory and Empirical Support." *ETS Research Report Series*, 1996(1), pp. 1 - 41.
- Holstrom, B. and Milgrom, P. (1991). "Multitask principle-agent analyses: incentive contracts, asset ownership and job design." *Journal of Law, Economics and Organization*, Issue 7, pp. 24 - 52.
- Jacob, B.A. and Levitt, S.D. (2003). "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *The Quarterly Journal of Economics*, pp. 843 - 877.
- Jacob, B.A. (2005). "Accountability, Incentives and Behaviour: the Impact of high-stakes testing in the Chicago Public Schools." *Journal of Public Economics*, Issue 89, pp. 761 - 796.
- Jacob, B.A. and Lefgren, L. (2003). "Are Idle Hands the Devil's Workshop? Incapacitation, Concentration, and Juvenile Crime." *The American Economic Review*, 93(5), pp. 1560 - 1577.
- Josephson Institute Centre for Youth Ethics. (2012). "The Ethics of American Youth 2012." [Online] Available at: <https://charactercounts.org/programs/reportcard/2012/installment-report-card-honesty-integrity.html> [Accessed 19 October 2015].
- Kremer, M., Chaudhury, N., Rogers, F.H., Muralidharan, K. and Hammer, J. (2005). "Teacher Absence in India: A Snapshot." *Journal of the European Economic Association*, 3(2 - 3), pp. 658 - 667.
- Lazear, E. (2003). "Teacher Incentives." *Swedish Economic Policy Review*, 10(2), pp. 179 - 214.
- Lieberman, E.S., Posner, D.N. and Tsai, L.L. (2014). Does information lead to more active citizenship? Evidence from an education intervention in rural Kenya. *World Development*, 60, pp. 69 - 83.
- Linn, R. (2004). "Chapter 4: Accountability Models." In: Furlman, S.H. and Elmore, R.F. ed. *Redesigning Accountability Systems for Education*. New York : Teachers College Press, pp. 73 - 95.
- Loveless, T., Costrell, R.M. and Cuban, L. (2005). "Test-based Accountability: The Promise and the Perils." *Brookings Papers on Education Policy*, Issue 8, pp. 7 - 45.
- Lucas, R.. (1988). "On the Mechanics of Economic Development." *Journal of Monetary Economics*, July, 22(1), pp. 3 - 42.
- Mankiw, N.G., Romer, D. and Weil, D. (1990). "A Contribution to the Empirics of Economic Growth." (No. w3541). *National Bureau of Economic Research*.

- McCabe, D.L., Trevino, L.K. and Butterfield, K.D. (2001). "Cheating in Academic Institutions: A Decade of Research." *Ethics & Behaviour*, 11(3), pp. 219 - 232.
- McNeil, L.M., Coppola, E., Radigan, J. and Heilig, J.V. (2008). "Avoidable Losses: High-stakes Accountability and the Dropout Crisis." *Archivos Analíticos de Políticas Educativas = Education Policy Analysis Archives*, 16(3), p. 1.
- Mhlanga, E. (2015). "5 300 matric candidates under investigation over mass copying." Department of Basic Education.
- Modisaotsile, B.M. (2012). "The Failing Standard of Basic Education in South Africa." *Africa Institute of South Africa*, Volume 72.
- Moloi, M.Q. and Chetty, M. (2010). "The SACMEQ III Project in South Africa: A Study of the Conditions of Schooling and the Quality of Education." Pretoria: Department of Education.
- Mulligan, C. (1999). "Galton versus the Human Capital Approach to Inheritance." *Journal of Political Economy*, 107(6), pp. 184 - 224.
- Muralidharan, K. and Sundararaman, V. (2010). "The Impact of Diagnostic Feedback to Teachers on Student Learning: Experimental Evidence from India." *The Economic Journal*, 120(546), pp. F187 - F203.
- Murnane, R.J., Willet, J.B., Duhaldeborde, Y. and Tyler, J.H. (2000). "How Important are the Cognitive Skills of Teenagers in Predicting Subsequent Earnings?" *Journal of Political Analysis and Management*, pp. 547 - 568.
- Nichols, S.L., Berliner, D.C. and Noddings, N. (2007). "Collateral Damage: How High-stakes Testing Corrupts America's Schools."
- Pratham, P. (2008). *Annual Status of Education Report (Rural), (2007) (No. id: 1802)*.
- Prew, M. (2009). "Community Involvement in School Development Modifying School Improvement Concepts to the Needs of South African Township Schools." *Educational Management Administration & Leadership*, 37(6), pp. 824 - 846.
- Shalem, Y. (2003). "Do we have a Theory of Change? Calling Change Models to Account." *Perspectives in Education: Assessment of Change in Education*, Volume Special Issue 1, 21.
- Smith, P. (1993). "Outcome-related Performance Indicators and Organisational Control in the Public Sector." *British Journal of Management*, Volume 4, pp. 135 - 151.
- Spaull, N. (2013). "South Africa's Education Crisis: The Quality of Education in South Africa 1994-2011." Johannesburg: Centre for Development & Enterprise.
- Springer, M. (2008). "The Influence of an NCLB Accountability Plan on the Distribution of Student Test Score Gains." *Economics of Education Review*, 27(5), pp. 556 - 563.
- Transparency International. (2014). "Transparency International Corruption Perceptions Index." Ernst & Young.

United Nations Decade of Education for Sustainable Development. (2002). "Background: Agenda 21, Chapter 36." [Online] . Available at: <http://www.gdrc.org/sustdev/un-desd/c36-a21.html> [Accessed 7 October 2015].

Vilakazi, N. and Gustafsson, M. (2014). "Report on 2013 ANA Data Consistency."

Volmink, J., Gardiner, M., Msimang, S., Nel, P., Moleta, A., Scholtz, G., and Prins, T. (2016). Report of the Ministerial Task Team appointed by Minister Angie Motshekga to investigate Allegations into the selling of posts of educators by members of Teachers Unions and Departmental Officials in Provincial Education Departments. Department of Basic Education.

Wollack, J.A. (1997). "A Nominal Response Model Approach for Detecting Answer Copying." *Applied Psychological Measurement*, 21(4), pp. 307 - 320.

Appendix

Question to Item Index

Grade 3 Mathematics

Question	Item	Marks
1	1	1
2	2	1
3	3	1
4	4	1
5	5	1
6	6	1
7.1	7	1
7.2	8	1
8	9	1
9.1	10	2
9.2	11	2
10.1	12	1
10.2	13	1
11	14	1
12	15	1
13	16	1
14	17	1
15	18	1
16	19	1
17	20	1
18.1	21	1
18.2	22	1
19	23	2
20	24	2
21	25	1
22.1	26	1
22.2	27	1
23.1	28	1
23.2	29	1
24.1	30	2
24.2	31	2
25	32	1
26	33	1
27	34	1
27	34	40

Grade 3 Home Language

Question	Item	Marks
1	1	1
2	2	1
3	3	1
4	4	1
5	5	1
6	6	1
7.1	7	1
7.2	8	1
8.1	9	2
8.2	10	2
9.1	11	2
9.2	12	2
9.3	13	2
10.1	14	1
10.2	15	1
10.3	16	1
11.1	17	2
11.2	18	2
11.3	19	2
12.1	20	1
12.2	21	1
13.1	22	1
13.2	23	1
13.3	24	1
14.1	25	1
14.2	26	1
15	27	6
15	27	40

Grade 6 Mathematics

Question	Item	Marks
1.1	1	1
1.2	2	1
1.3	3	1
1.4	4	1
1.5	5	1
1.6	6	1
1.7	7	1
1.8	8	1
1.9	9	1
1.10	10	1
2	11	1
3	12	1
4	13	1
5.1	14	2
5.2	15	2
5.3	16	4
5.4	17	3
5.5	18	2
5.6	19	2
5.7	20	3
6	21	1
7	22	1
8	23	2
9	24	1
10	25	1
11	26	3
12	27	3
13	28	1
14	29	1
15	30	2
16	31	2
17.1	32	1
17.2	33	1
18	34	2
19	35	3
20	36	1
21	37	2
22.1	38	1
22.2	39	1
23.1	40	1

23.2	41	1
24	42	2
25	43	1
26	44	2
27.1	45	1
27.2	46	1
28	47	1
29	48	2
30.1	49	1
30.2	50	1
30	50	75

Grade 6 First Additional Language

Question	Item	Marks
1	1	1
2	2	1
3	3	1
4	4	1
5	5	1
6	6	2
7	7	2
8	8	2
9	9	1
10	10	1
11	11	1
12	12	1
13	13	1
14	14	1
15	15	2
16	16	1
17	17	2
18	18	1
19	19	1
20	20	2
21	21	1
22	22	1
23.1	23	2
23.2	24	3
24.1	25	1
24.2	26	1
25	27	2
26	28	1
27.1	29	1
27.2	30	1
28	31	1
29.1	32	1
29.2	33	1
30.1	34	1
30.2	35	1
31.1	36	1
31.2	37	1
32	38	1
33	39	1
34	40	1

35	41	1
36.1	42	1
36.2	43	1
37	44	6
38.1	45	5
38.2	46	5
38.3	47	5
38	47	75