# Has there been score (grade) inflation in the South African National Senior Certificate from 2009 to 2013?

Robert Prince and William Msemburi

University of Cape Town

Suggestion for citation:

Prince, R and Msemburi, W. 2014. Has there been score (grade) inflation in the South African National Senior Certificate from 2009 to 2013? Working Paper: Centre for Educational Testing for Access and Placement (CETAP), Centre for Higher Educational Development (CHED), University of Cape Town (UCT).

Ideally school leaving examination results signal to potential employers and higher education institutions the quality or ability of school leavers. This signal is particularly important in countries where unemployment is high and in situations such as when considering admissions into higher education where places are highly sought after. If the examination ranks students well and this is the main consideration then score inflation is less of a concern than it might be otherwise. In South Africa there is a national school leaving examination which is largely the basis on which higher education applicants are made offers of enrolment. Higher education dropout rates are also high which makes it important that the signal be both reliable and consistent so that admission decisions can be optimal. Grade inflation, whereby scores increase on an examination over time without an improvement in underlying ability, is one of the factors that makes it hard to interpret these examination scores for higher education admissions purposes. The purpose of this paper is to

1. Develop a methodology for determining score inflation in the South African National Senior Certificate examination;

2. To use the score inflation methodology to evaluate the evidence on whether there is score inflation in the South African national school leaving examinations;

3. Does score inflation vary across provinces.

Keywords

Grade inflation; Score inflation; High School Average Score;

## Introduction

One of the many challenges facing higher education institutions in South Africa is having a means to complement information on the school exit outcomes with information about student preparedness for the expectations of university study. School exit outcomes differ from University preparedness in that they provide information on what candidates have learned but not the extent to which they are able to apply what they have learnt at school in more independent teaching and learning environments like those experienced at university.

The National Senior Certificate (NSC) provides information on school exit outcomes while the National Benchmark Tests (NBTs) provide information on academic readiness for higher education. This difference in the information provided by the NSC and the NBT has created some of the causes of misunderstandings of the role of the NBTs in the South African higher and basic education landscape.

Yeld states this misunderstanding succinctly in 2010 *"The misunderstandings evident in the public discourse around both the NSC and NBT results point to an urgent need for a clarification of the nature of each of these assessment regimes, the approaches employed in their development and the role each has to play in the system. Broadly speaking, the major differences are in the purposes of the tests, their developmental approaches and their approaches to standardisation", p4.).*

The aim of this paper is to contribute to the debate on the relationship between the NSCs and NBTs by especially examining the extent to which the NBTs, which are standardized tests, can be used to evaluate whether current trends in NSC results are indicative of score (grade) inflation.

In recent years, there is a perception that the NSC scores obtained by students do not match their level of performance at university first year level. South African universities have for a long time been increasingly concerned about whether the matriculation certificate is the best predictor for academic success at tertiary level, while increasingly the perception that student performance at tertiary levels seems poorer than would be suggested by their matriculation results. This situation suggests the possibility that a rise in NSC grades may be occurring without the corresponding rise in student ability or performance, a situation which reflects a phenomenon called "grade inflation" and which I wish to call "score inflation."

The National Senior Certificate (NSC) examined for the first time at the end of 2008, created uncertainties among universities and produced an unusually high number of learners who qualified for university admission based on their NSC results. Yet by 2009, the large abnormal influx of first year students left several tertiary institutions with higher than normal failure rates. How can it be that such good NSC results were associated with very poor first year university performance unless the NSC grades inflated student ability?

In this paper we use the NBT results which are criterion-referenced and standardised indicators of ability to investigate if there is evidence of score inflation in the norm referenced NSC examination results over the period 2009 to 2013. A national sample of students is used from 2009 to 2013 which includes only those students who have both NSC scores and NBT scores.

Before presenting the data, we review briefly the differences in measurement between the NSCs and NBTs.

## Differences in approaches to measurement

The NBTs are all based on well researched academically sound test specifications, uses modern test theories to determine test scores and allows for the comparison of scores achieved during different test administrations including different years. It also uses criterion-referenced benchmarks set through international best practice standard setting methods to place candidates in proficiency levels which describes both their preparedness for the demands of higher education and the extent to which higher education's curricula should be responsive to the preparedness of the candidates they do admit.

The aspects of the test specifications that are in the public domain is in the book Access and Entry level Benchmarks edited by Hanlie Griesel in 2006. Common items and the 3 parameter Item Response Theory (IRT) model is used for scoring and equating, the process by which scores on different tests are made comparable. The standard-setting method that was employed to determine the benchmarks is the modified Angoff method. This standard-setting process uses expert panels made up of lecturers that teach first year courses in higher education and essentially is tasked with answering the question for each item in a test "what is the probability of candidates at the borderline of the proficiency categories Basic-Intermediate and Intermediate-Proficient successfully answering a given question correctly?" and then aggregating this up for the entire test. The NBTs test three domains of Academic Literacy (AL); Quantitative Literacy (QL) and Mathematics (MAT).

Academic Literacy (AL) assesses a students' capacity to engage successfully with the language demands of academic study in the medium of instruction.

Quantitative Literacy (QL): assesses a students' ability to manage situations or solve problems of a quantitative nature in real contexts relevant to quantitative disciplines in higher education.

Mathematics (MAT): assesses a students' manifest ability related to mathematical concepts formally part of the School Mathematics curriculum relevant to Mathematics and mathematical disciplines in higher education (Prince, 2013).

The NBTs have test proficiency levels that are described below. These describe the expected performances for each benchmark. The current benchmark levels are displayed in the tables below.

Table 1.1: National Benchmark Test performance standards and their interpretations

| | | |
|---|---|---|
| Proficient | 100 | Test performance suggests that future academic performance will not be adversely affected (students may pass or fail at university, but this is highly unlikely to be attributable to strengths or weaknesses in the domains tested). If admitted, students may be placed into regular programmes of study.<br>AL (65%) new [64%]; QL old (66%) new [70%) MAT old (62%) new [68%] |

| | | |
|---|---|---|
| Intermediate | | The challenges identified are such that it is predicted that academic progress will be adversely affected. If admitted, students' educational needs should be met as deemed appropriate by the institution (e.g. extended or augmented programmes, special skills provision).<br>AL old(42%) new [38%]; QL old(38%) new [38]; MAT old (34%) new [35%] |
| Basic | 0 | Test performance reveals serious learning challenges: it is predicted that students will not cope with degree-level study without extensive and long-term support, perhaps best provided through bridging programmes (i.e. non-credit preparatory courses, special skills provision) or FET provision. Institutions admitting students performing at this level would need to provide such support themselves. |

*Adapted from Yeld, N (2010) Prince, R., (2013)*

In addition, it has been found productive to divide the intermediate performance bands into Upper and Lower as shown in table 1.2. It is important to note that this division was not done through the standard-setting exercise but rather through taking the interval mean values.

Table 1.2: NBT degree Intermediate benchmarks and how they should be interpreted

| | Upper intermediate | Assessment of need | Lower Intermediate | Assessment of need |
|---|---|---|---|---|
| AL | 53-64 [51-63] | Students are likely to need complementary support(additional tutorials, workshops, augmented courses, language intensive work) | 45-52 [38-50] | Students need to be placed in an extended programme |
| QL | 52-65 [54-69] | | 38-51 [38-53] | |
| MAT | 48-61 [52-67] | | 34-47 [35-51] | |

**Norm referenced tests**

The NSC subject scores are norm referenced. Norm-referenced tests (NRTs) are made to compare test takers to each other. NRTs are designed to sort and rank students usually "on a curve," not to see if they met a standard or criterion. Norm-referenced tests compare an individual's performance to the performances of a group, called the "norm group." So the NSC cannot be used to assess whether candidates met a certain standard in a domain. The NSC has part of the final scores made up of the course mark and then the scores are 'standardised' to the 5-year rolling average score for each subject. So while a candidate may perform very well compared to the norm, they may still fail to meet a certain standard in the domain being tested.

NSC subjects are categorised according an achievement scale of 1 to 7 with descriptors as in the table below. These are not benchmarks, but rather descriptive categories of what a percentage mark means in terms of a candidate's achievement.

The Department of Basic Education (DBE) approved the following achievement scale for NSC subjects (DBE, 2011 p.9)

**Table 1.3: NSC scale of achievement Grades 10-12**

| Rating Code | Description | Score |
|---|---|---|

| | | |
|---|---|---|
| 7 | *Outstanding achievement* | *80-100* |
| 6 | *Meritorious achievement* | *70-79* |
| 5 | *Substantial achievement* | *60-69* |
| 4 | *Adequate achievement* | *50-59* |
| 3 | *Moderate achievement* | *40-49* |
| 2 | Elementary achievement | 30-39 |
| 1 | Not achieved | 0-29 |

On completing the NSCs, a candidate can qualify for certificate, diploma or degree study. The following criteria for entry into Higher Certificate, Diploma or Degree Study are used in South Africa.

**Table 1.4 Criteria for Higher certificate, Diploma and Degree study**

| **Qualification** | **Minimum Entry requirement** |
|---|---|
| Higher Certificate | Pass NSC with at least rating of 2 for the Language of Learning and Teaching (LOLT) of higher education institution |
| Diploma | Pass NSC with an achievement rating of 3 (40-49%) or better in four subjects. At least rating of 2 for the Language of Learning and Teaching (LOLT) of the higher education institution. |
| Bachelor Degree | Pass NSC with an achievement rating of 4 (50-59%) or better in four subjects from the designated list. At least rating of 2 for the Language of Learning and teaching (LOLT) of the higher education institution. |
| | |

Source**:** Department of Basic Education, 2012.p.11.

**Test development of the NSC**

The Department of Basic Education moved from provincial to national test questions in 2008. The department also employs some form of standard setting (DBE *op.cit*. p.19), with the question papers set by a panel of 3-5 members representative of as many provinces as possible and with the right subject expertise. One internal moderator is also employed during the examination setting exercise. Eighty-four external moderators from Umalusi were used to evaluate and approve question papers in November 2012 and March 2013 NSC examinations concurrently to ensure comparable standards. The DBE also engages in language simplification by checking correlations between Afrikaans and English versions of the question papers.

This brief overview of differences in measurement is important in order for this report to emphasise that score inflation is not a critique of DBE's examination procedures but rather a way in which we can use the standardised criterion-referenced test such as the NBTs and a norm referenced test such as the NSCs to refine the assessment of student performance.

**Data Sources for this paper**

The data for this study included high school candidates for 2009 and 2013 who took the NBT test in their twelfth grade of high school. High School scores were calculated by using the scores for the highest six subjects excluding Life Orientation. NBT data were obtained from the NBTP (national

benchmark test project) which is housed at UCT, while the NSC data are obtained from DBE database. The total numbers analysed by year and for each Department of Basic Education (DBE) province or Independent Examinations Board (IEB) is shown in Table 1.5 below. The figures reflect the number of candidates who wrote both the NSCs and NBTs in 2009, 2010, 2011, 2012 and 2013. The total number has increased by 16,453 writers within a five year period. An encouraging fact is that all the provinces show an increase, with the exception of the Free State.

*Table 1.5: NSC and NBT dataset for NBT writers nationally (N = 182,156), 2009-2013*

| Province and EB | NSC Exam Year | | | | | Total |
|---|---|---|---|---|---|---|
| | 2009 | 2010 | 2011 | 2012 | 2013 | |
| Eastern Cape Province (EC) | 1,966 | 3,055 | 3,942 | 4,134 | 4,001 | 17,098 |
| Free State Province (FS) | 2,200 | 1,688 | 2,344 | 2,187 | 2,084 | 10,503 |
| Gauteng Province (GT) | 7,844 | 9,113 | 8,898 | 9,287 | 10,202 | 45,344 |
| Independent Examinations Board (IEB) | 2,568 | 2,084 | 3,555 | 4,589 | 4,748 | 17,544 |
| KwaZulu natal Province (KN) | 3,025 | 3,477 | 4,075 | 4,611 | 6,280 | 21,468 |
| Mpumalanga Province (MP) | 1,276 | 1,714 | 2,016 | 1,886 | 1,983 | 8,875 |
| Northern Cape Province (NC) | 395 | 489 | 559 | 589 | 635 | 2,667 |
| Northern Province (NP) | 3,215 | 3,279 | 3,706 | 4,618 | 5,521 | 20,339 |
| North West Province (NW) | 964 | 1,144 | 1,101 | 1,206 | 1,285 | 5,700 |
| Western Cape Province (WC) | 5,339 | 5,155 | 5,411 | 8,207 | 8,506 | 32,618 |
| Total | 28,792 | 31,198 | 35,607 | 41,314 | 45,245 | 182,156 |

## Methodology

In evaluating score (grade) inflation, analysis began with calculating a composite measure for both the NSC scores and for the NBT AQL. The three composite measures are:-

**NSC600 -** Sum of the top 6 NSC subject results (out of 600) excluding life orientation for each candidate.

- **AQL -** Sum of NBT Academic Literacy (AL) and Quantitative Literacy (QL) scores (out of 200)
- **AQLMAT -** Sum of the NBT AL, NBT QL and NBT Mathematics (MAT) scores (out of 300)

Distributions of frequencies by score, measure and year are shown below.

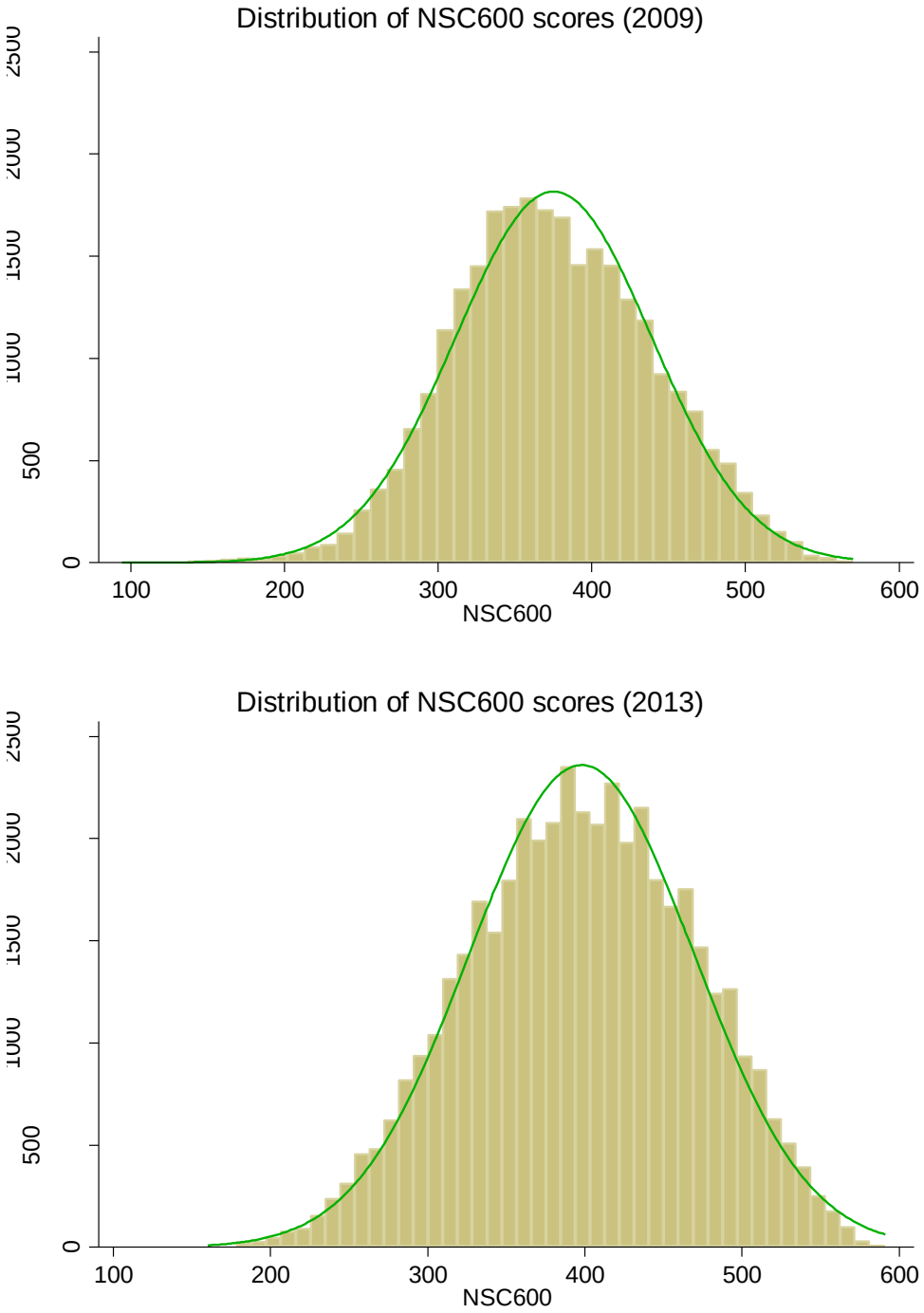



*Figure 1: Distribution of NSC600 scores, 2009 and 2013*

The distributions of NSC600 composite score for 2009 and 2013 are both normal, though there are indications that the frequencies with higher scores increased in 2013.

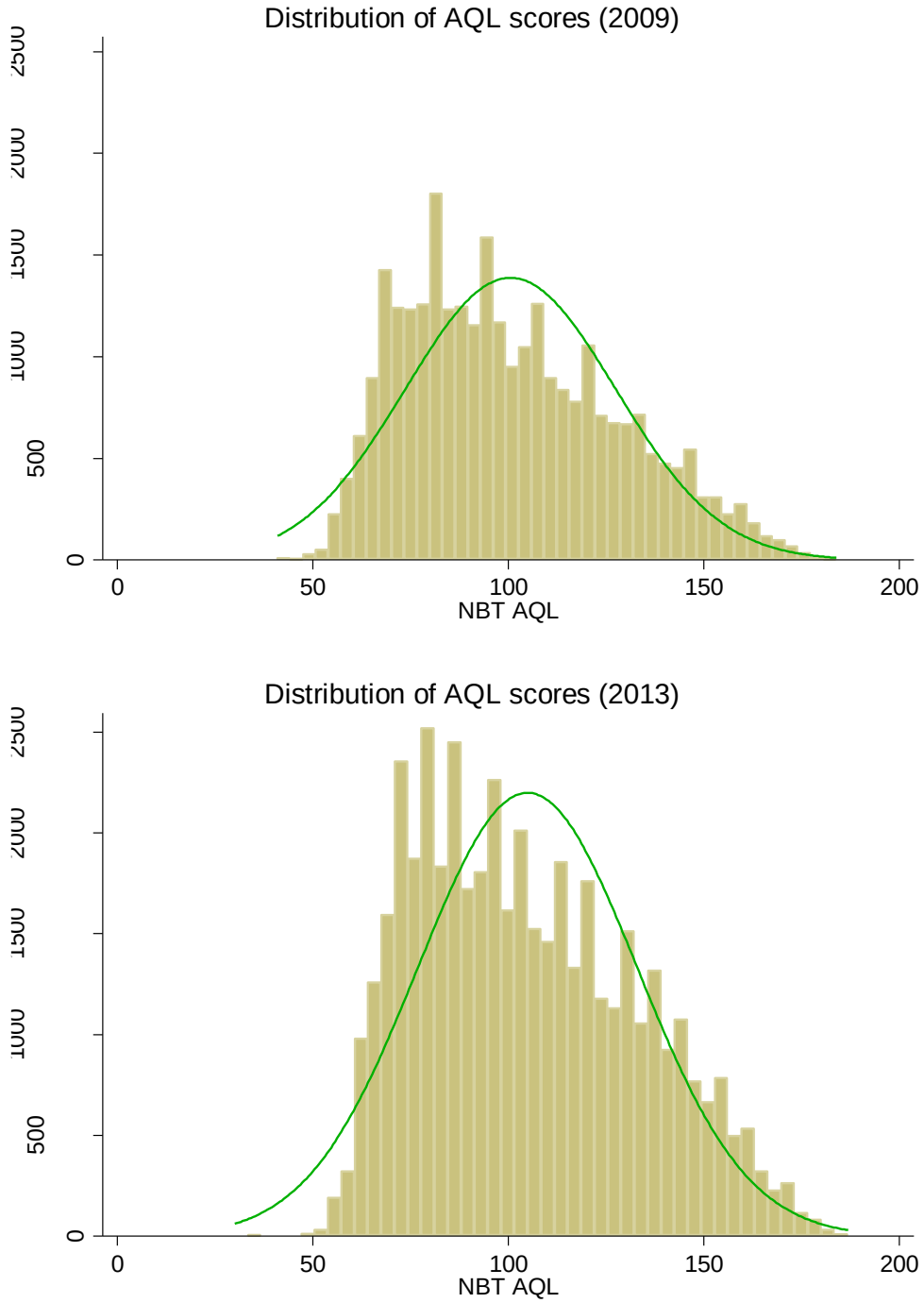



*Figure 2: Distribution of NBT AQL scores, 2009 and 2013*

Both the distribution of AQL and AQLMAT scores shows a higher frequency of those obtaining lower scores in 2013 than in 2009. Although this change is slight, it is unexpected, given the distribution for NSC 600 scores.

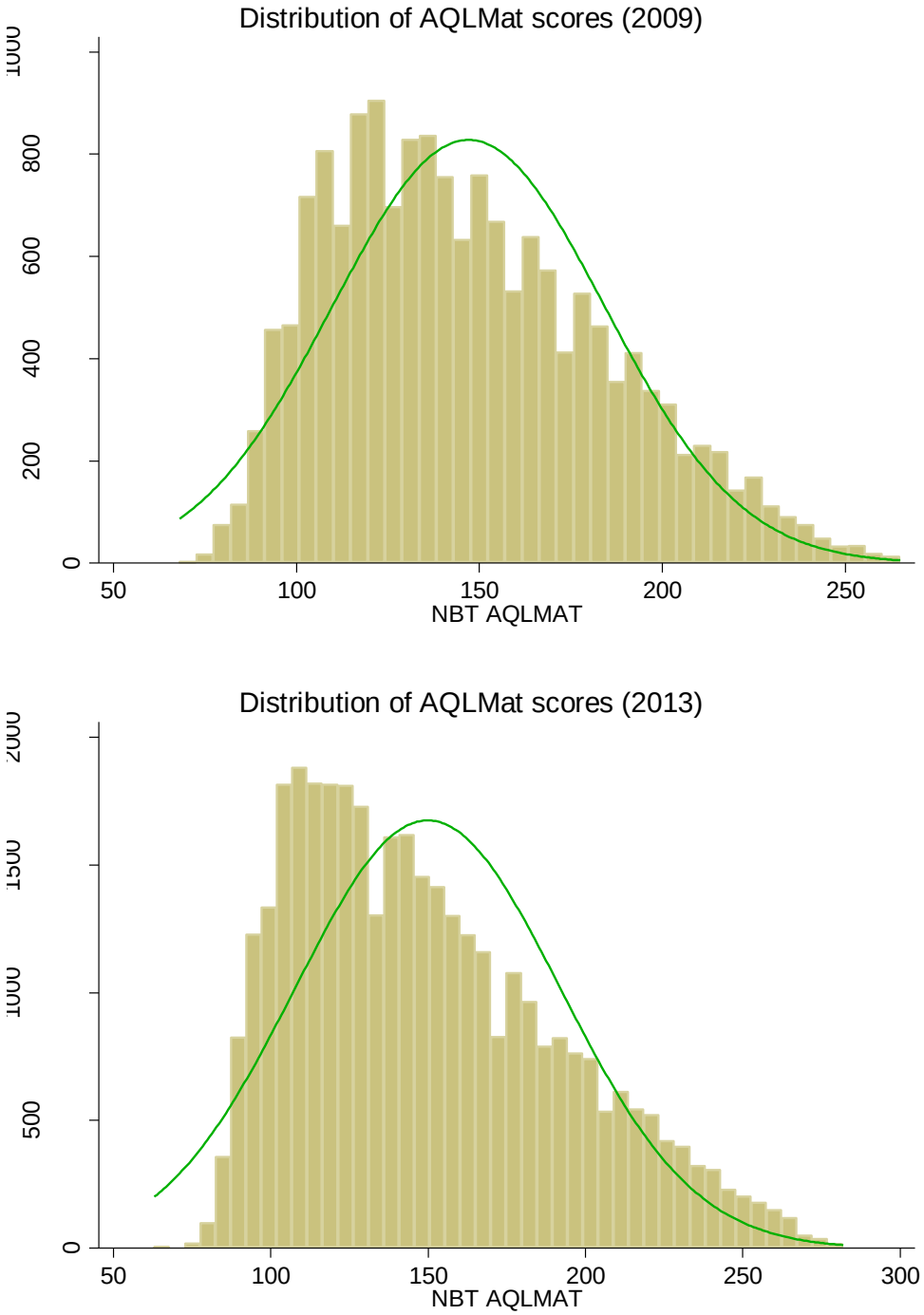



*Figure 3: Distribution of NBT AQLMAT, 2009 and 2013*

## Correlations between AQL, AQLMAT and NSC600 for 2009 and 2013

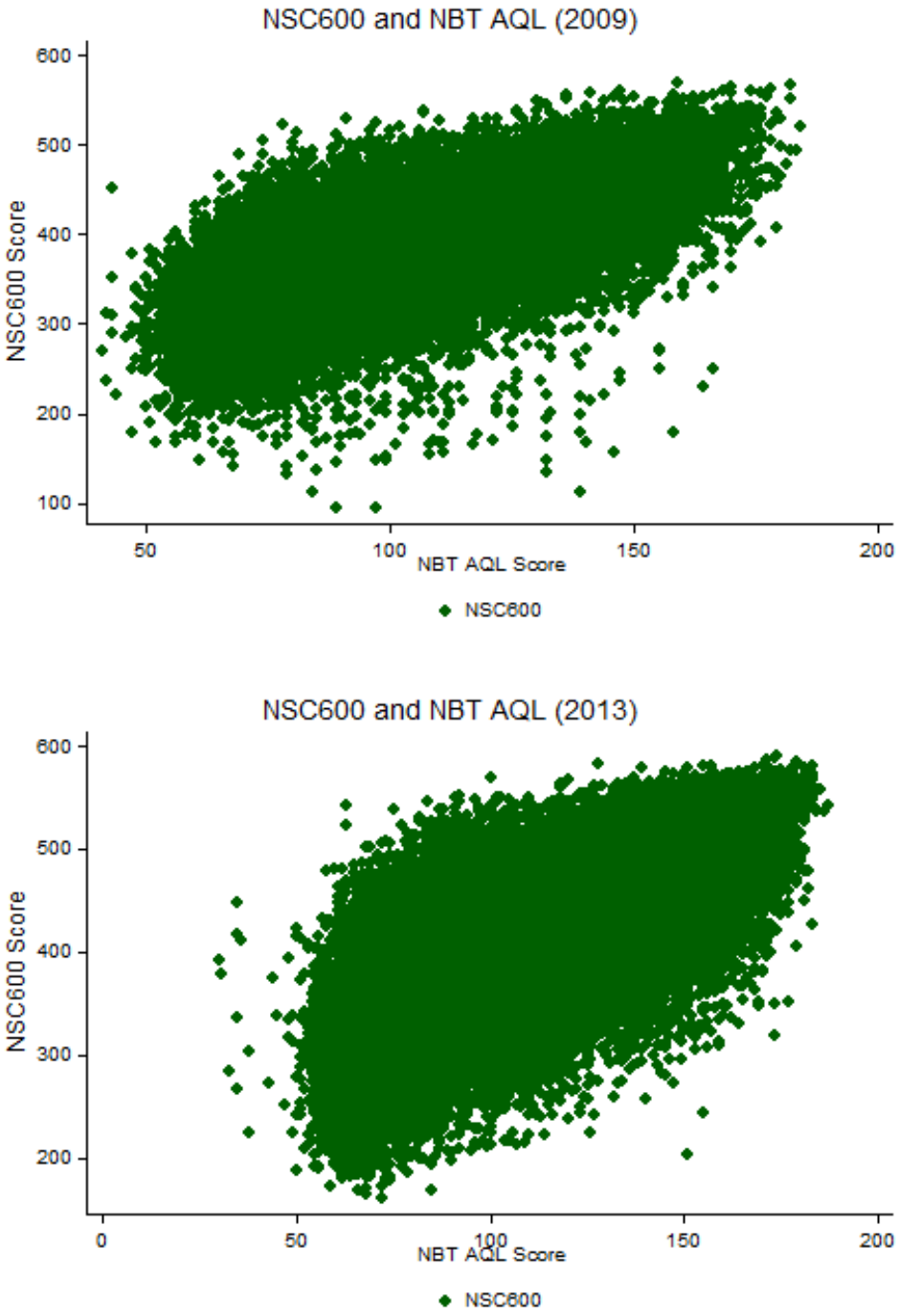
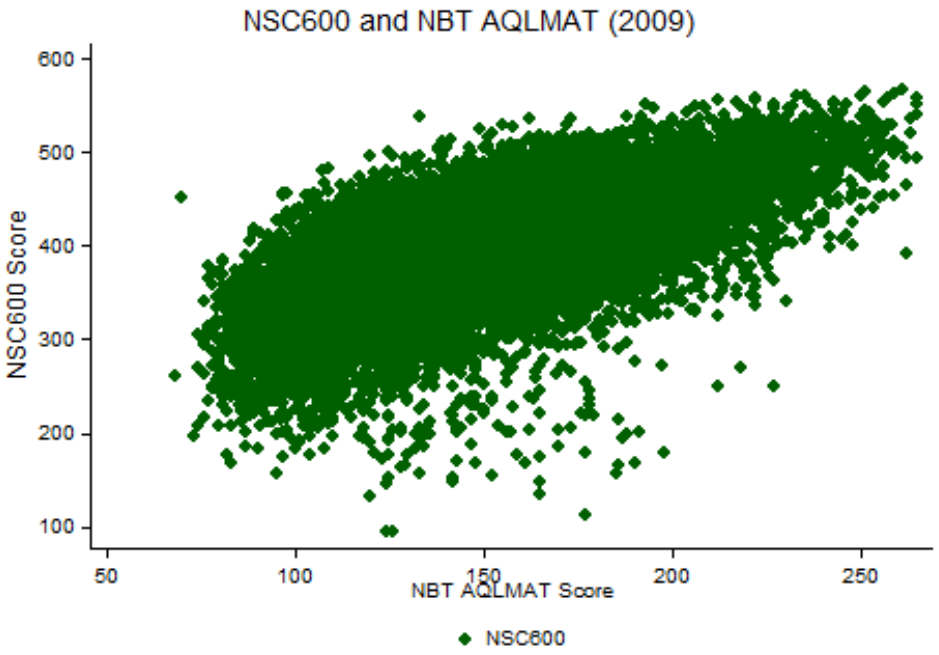



*Figure 4: NSC600 score by NBT AQL performance, 2009 and 2013*

The correlations between NSC600 and NBT AQL provide an interesting pattern. In 2009 there were fewer candidates who scored high NSC 600 scores and low AQL scores. In 2013 there are more candidates with high NSC 600 scores and low NBT scores.



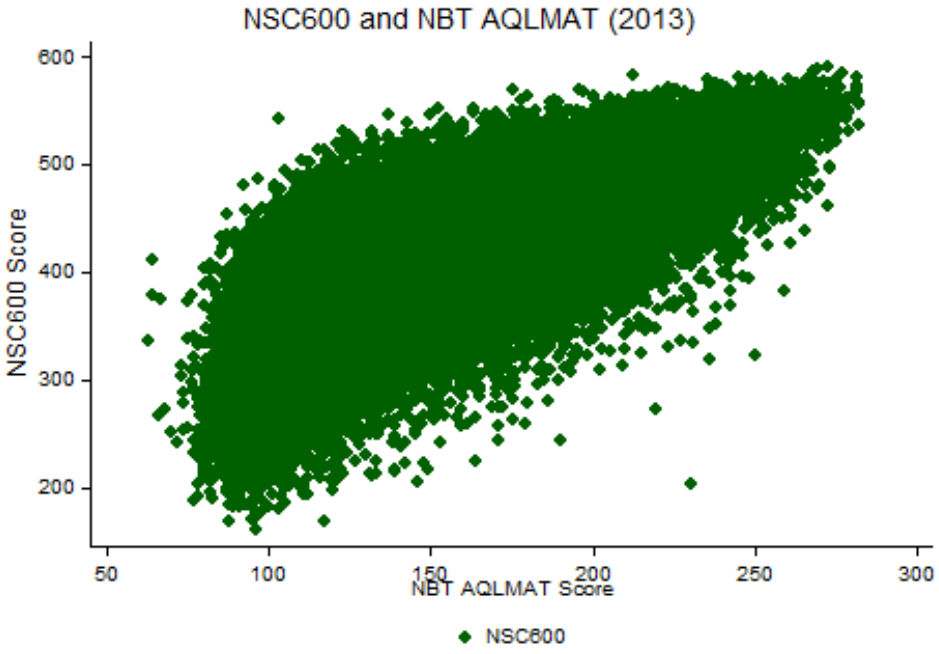




*Figure 5: NSC600 score by NBT AQLMAT performance, 2009 and 2013*

Since not everyone wrote the NBT MAT test and the correlations between AQL and NSC subjects were consistent during both 2009 and 2013, we used AQL as the conditional measure. We might have used any of AQL, AL, QL or MAT or even AQLMAT.

It is worth explaining the patterns of these correlations. AQL is strongly related to languages both English Home Language (ENHN) and English First Additional Language (ENFN). This makes sense since the AL domain is cognitively closest to the NSC language subjects than to the other NSC subjects.. The correlations between MAT and the NSC subjects Mathematics (MTHN) and Physical Science (PSCN) are high since they are cognitively close.

*Table 1 : Correlation matrix of 2009 NSC and NBT scores (* shows 0.01 significance level)*

| | AL | QL | MAT | AQL | AQLMAT |
|---|---|---|---|---|---|
| **ENFN** | 0.6113* | 0.4113* | 0.3390* | 0.5609* | 0.5148* |
| **LFSN** | 0.6125* | 0.5760* | 0.5673* | 0.6441* | 0.6525* |
| **MTHN** | 0.4677* | 0.5624* | 0.6796* | 0.5598* | 0.6181* |
| **MTLN** | 0.5576* | 0.5267* | -0.1449* | 0.6093* | 0.2010* |
| **PSCN** | 0.5441* | 0.5829* | 0.6871* | 0.6073* | 0.6634* |
| **GEON** | 0.6480* | 0.5733* | 0.5592* | 0.6604* | 0.6662* |
| **ENHN** | 0.6950* | 0.5212* | 0.5087* | 0.6510* | 0.6229* |

*Table 2 : Correlation matrix of 2010 NSC and NBT scores (* shows 0.01 significance level)*

| | AL | QL | MAT | AQL | AQLMAT |
|---|---|---|---|---|---|
| **ENFN** | 0.6417* | 0.5110* | 0.3601* | 0.6189* | 0.5948* |
| **LFSN** | 0.5844* | 0.5954* | 0.5406* | 0.6321* | 0.6511* |
| **MTHN** | 0.4654* | 0.6050* | 0.6611* | 0.5768* | 0.6474* |
| **MTLN** | 0.5167* | 0.5591* | -0.1078* | 0.5892* | 0.2480* |
| **PSCN** | 0.5078* | 0.5865* | 0.6394* | 0.5860* | 0.6421* |
| **GEON** | 0.6093* | 0.5951* | 0.5276* | 0.6471* | 0.6653* |
| **ENHN** | 0.7180* | 0.6064* | 0.5226* | 0.7076* | 0.6754* |

*Table 3: Correlation matrix of 2011 NSC and NBT scores (* shows 0.01 significance level)*

| | AL | QL | MAT | AQL | AQLMAT |
|---|---|---|---|---|---|
| **ENFN** | 0.7016* | 0.5737* | 0.4409* | 0.6822* | 0.6565* |
| **LFSN** | 0.6345* | 0.6235* | 0.6118* | 0.6711* | 0.7053* |
| **MTHN** | 0.4796* | 0.6235* | 0.7554* | 0.5929* | 0.6947* |
| **MTLN** | 0.5673* | 0.5905* | -0.1922* | 0.6309* | 0.2385* |
| **PSCN** | 0.4958* | 0.5747* | 0.6831* | 0.5707* | 0.6483* |
| **GEON** | 0.6379* | 0.6079* | 0.5375* | 0.6651* | 0.6773* |
| **ENHN** | 0.7211* | 0.5761* | 0.4961* | 0.6881* | 0.6423* |

*Table 4: Correlation matrix of 2012 NSC and NBT scores (* shows 0.01 significance level)*

| | AL | QL | MAT | AQL | AQLMAT |
|---|---|---|---|---|---|
| ENFN | 0.6963* | 0.5502* | 0.4294* | 0.6734* | 0.6364* |
| LFSN | 0.6267* | 0.6252* | 0.6225* | 0.6736* | 0.7093* |
| MTHN | 0.4505* | 0.6037* | 0.7486* | 0.5721* | 0.6876* |
| MTLN | 0.6199* | 0.6457* | -0.1892* | 0.6953* | 0.3108* |
| PSCN | 0.4641* | 0.5571* | 0.6879* | 0.5500* | 0.6427* |
| GEON | 0.6290* | 0.6067* | 0.5667* | 0.6642* | 0.6838* |
| ENHN | 0.7179* | 0.5728* | 0.5330* | 0.6923* | 0.6598* |

*Table 5: Correlation matrix of 2013 NSC and NBT scores (* shows 0.01 significance level)*

| | AL | QL | MAT | AQL | AQLMAT |
|---|---|---|---|---|---|
| ENFN | 0.7000* | 0.5414* | 0.4491* | 0.6699* | 0.6173* |
| LFSN | 0.6257* | 0.6182* | 0.6490* | 0.6711* | 0.7329* |
| MTHN | 0.4418* | 0.5919* | 0.7777* | 0.5616* | 0.7071* |
| MTLN | 0.6551* | 0.6994* | 0.3679* | 0.7426* | 0.6756* |
| PSCN | 0.4676* | 0.5552* | 0.7392* | 0.5511* | 0.6793* |
| GEON | 0.5814* | 0.5488* | 0.5736* | 0.6082* | 0.6659* |
| ENHN | 0.7235* | 0.5796* | 0.5629* | 0.6983* | 0.6667* |

## Grade Inflation- the method and evidence

Taking the NBT AQL as a standard measure, we expect the increases in NSC600 scores to mirror that of the AQL scores assuming that the increase in NSC scores are indicative of increase in student ability. However, when we plot the NBT AQL scores versus NSC600 scores, as in figure 6, it shows that a mean NBT score of 150 for example was associated with an NSC600 mean score of around 440 in 2009; while the same NBT AQL mean score was associated with a NSC600 mean score of close to 460 in 2013. This is strongly suggestive of NSC grade inflation.

The same pattern is also reflected when examining the relationship between NSC600 and AQLMAT.

In looking at the relationship between NBT MAT and PSCN, evidence of grade inflation is quite strong, especially among middle performing candidates. However, the relationship between NBT MAT and MTHN is the only one where we found both inflation (at the low end) and inflation (at the high end). The mean MAT score of 50 is associated with an NSC PSCN means score of 60 in 2009, and of 70 in 2013. This increase is indicative of grade inflation. Figure 11 is interesting. Though there seems to be significant grade inflation among the poorly performing candidates, there is evidence of less inflation among the high scoring candidates.

The relationship between NBT AL and ENFN shows clear and sustained grade inflation. This suggests that either the EFFN is being continuously marked generously to inflate student grades or the examination has been set to cater for students with low ability. Also for candidates with the same AL

scores candidates presenting ENFN consistently score approximately 10 percentage points higher than their ENHN counter parts.

## Five year trends of NSC600 using AQL as a stable measures

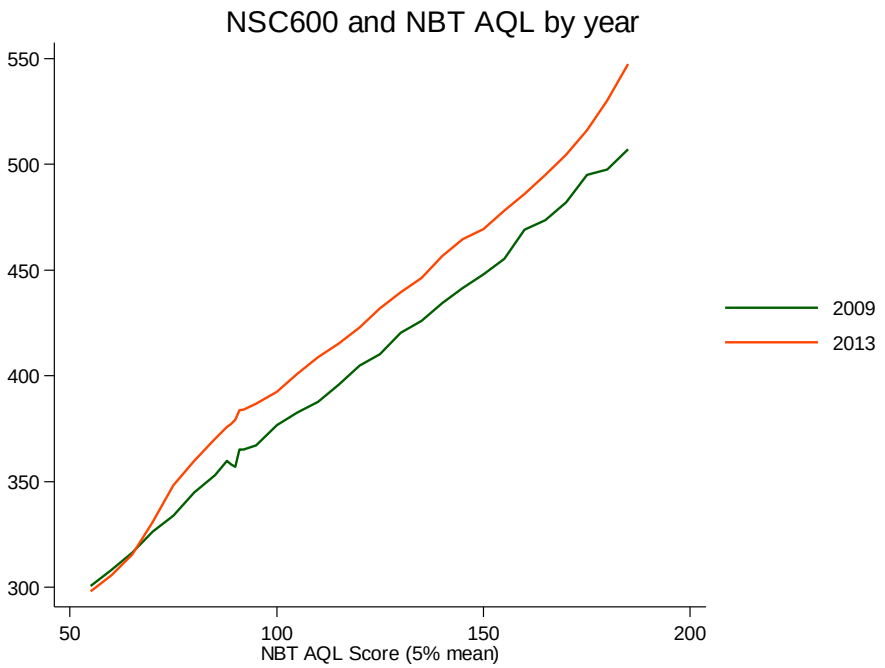



*Figure 6: Mean NSC600 by mean NBT AQL performance, 2009 and 2013*

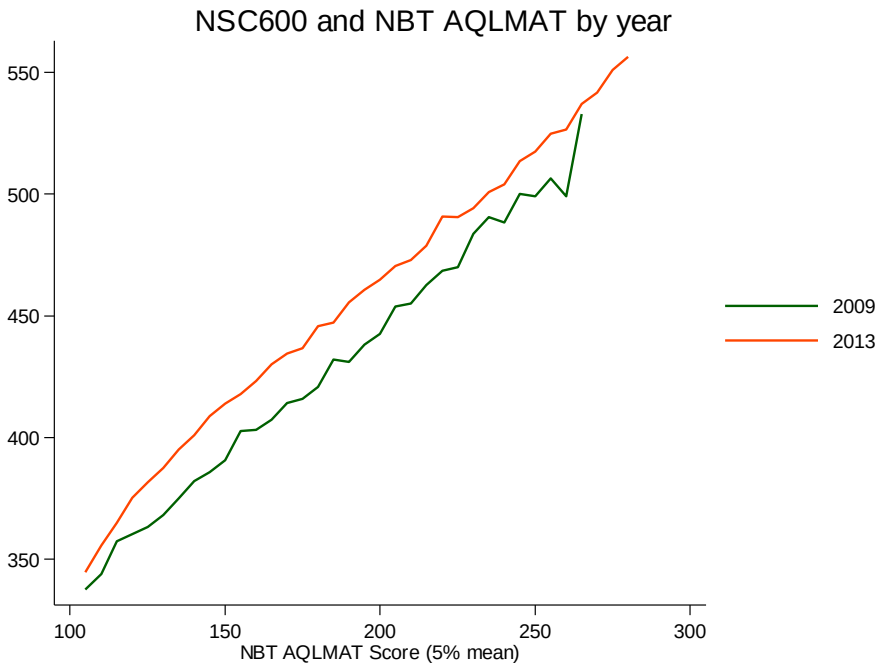


*Figure 7: Mean NSC600 by mean NBT AQLMAT performance, 2009 and 2013*

## Five year trends of NSC subjects using NBT subjects as stable measures

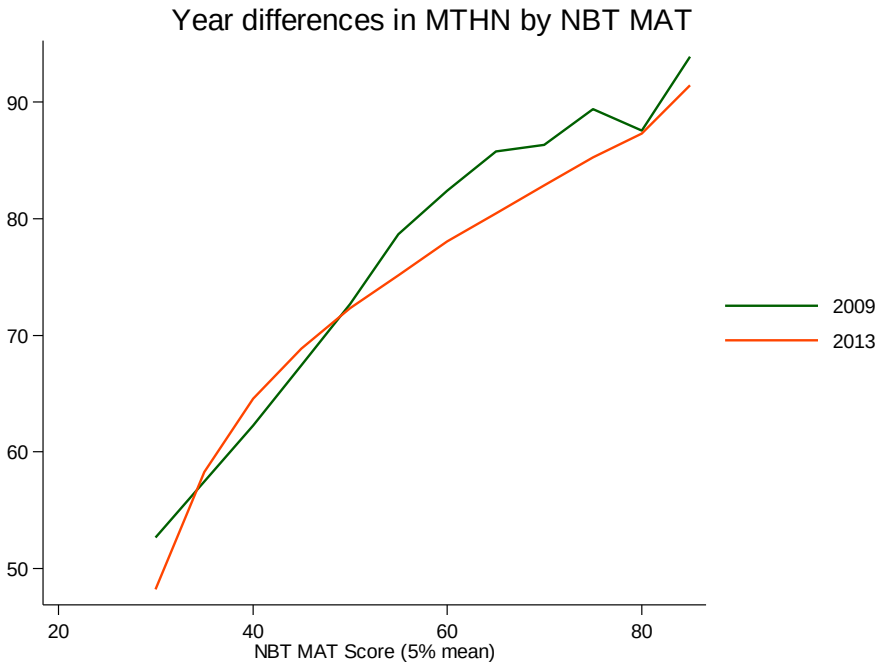



*Figure 8: Mean MTHN by mean NBT MAT performance, 2009 and 2013*

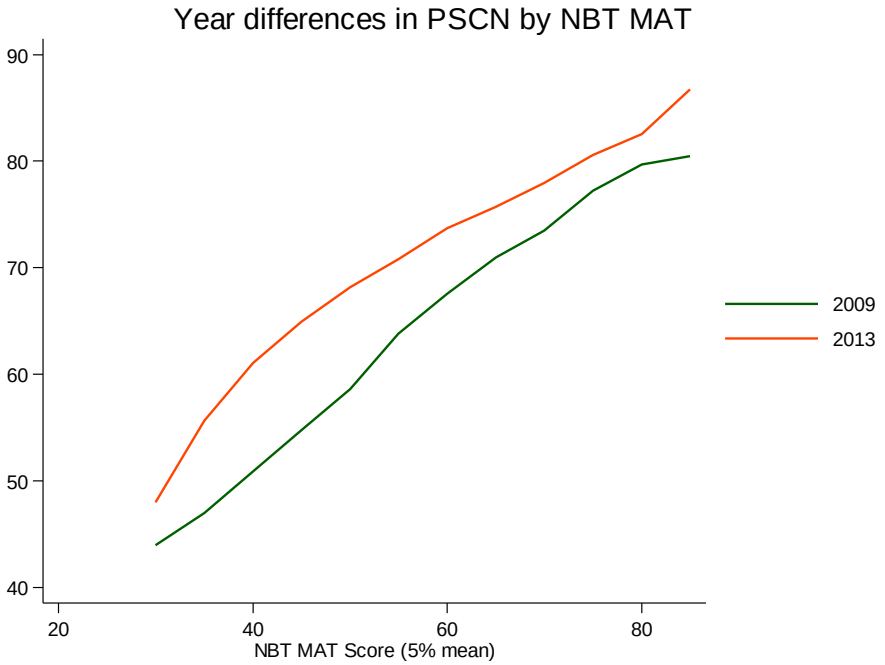



*Figure 9: Mean PSCN by mean NBT MAT performance, 2009 and 2013*

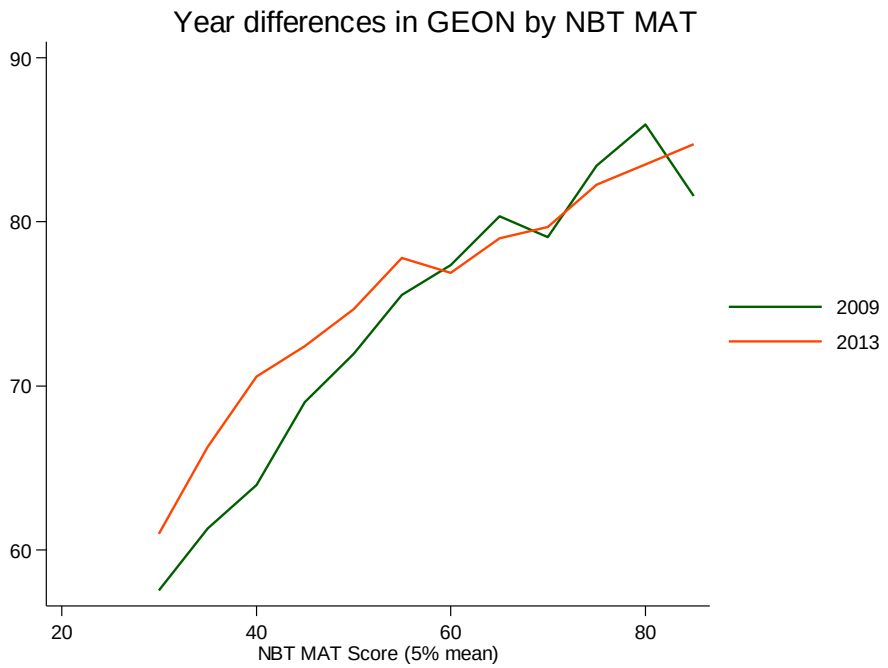



*Figure 10: Mean GEON by mean NBT MAT performance, 2009 and 2013*

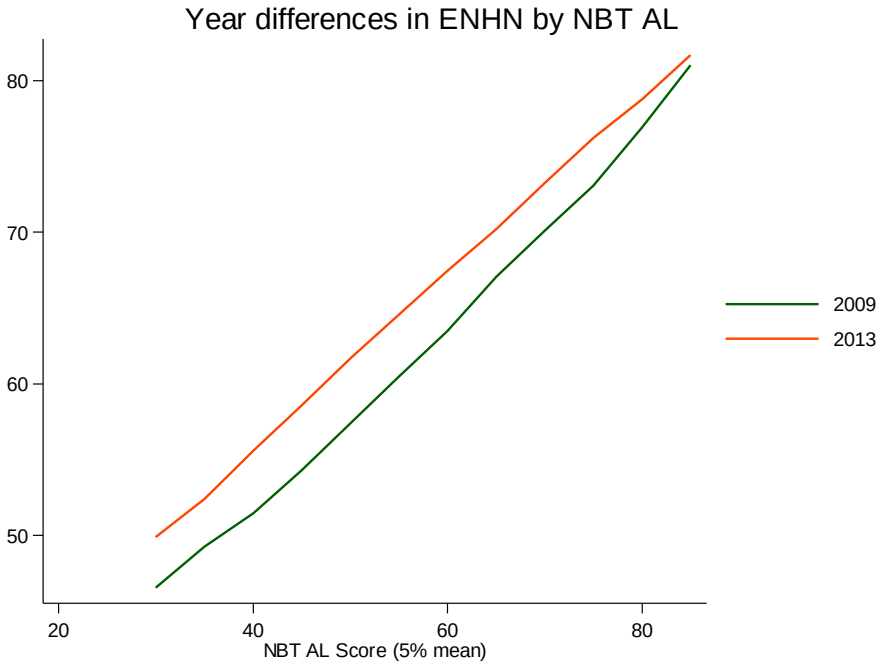



*Figure 11: Mean ENHN by mean NBT AL performance, 2009 and 2013*

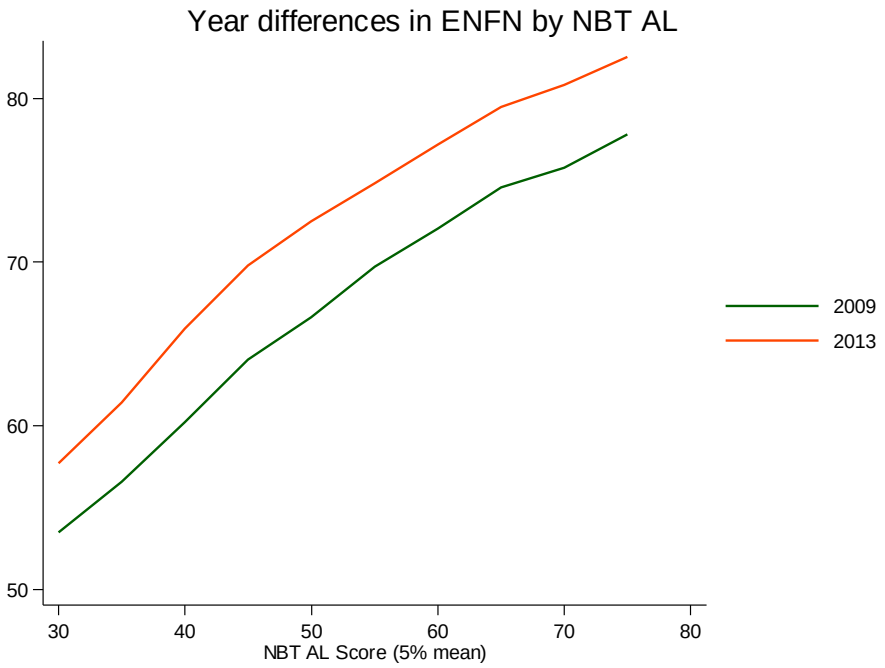



*Figure 12: Mean ENFN by mean NBT AL performance, 2009 and 2013*

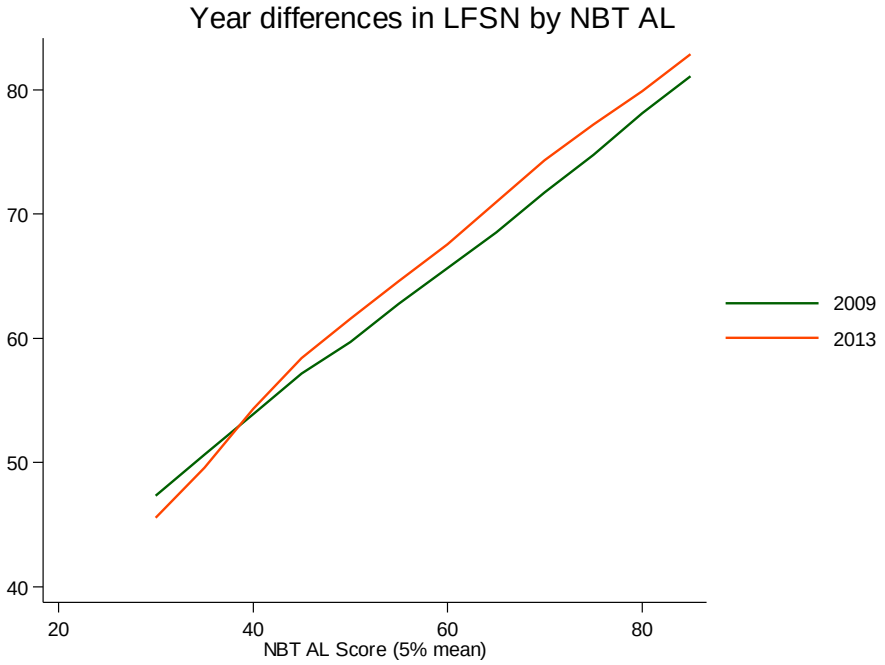



*Figure 13: Mean LSFN by mean NBT AL performance, 2009 and 2013*

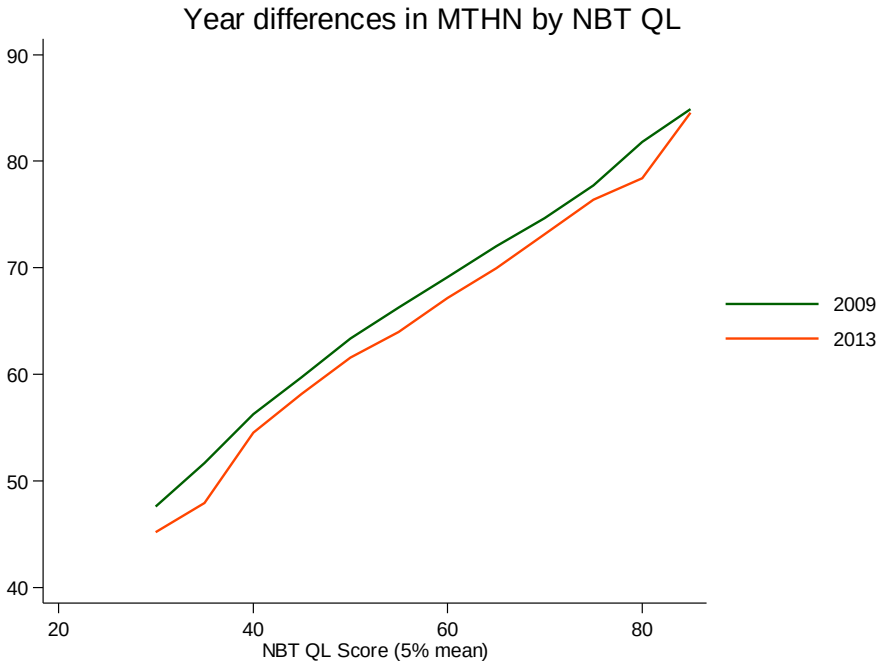



*Figure 14: Mean MTHN by mean NBT QL performance, 2009 and 2013*

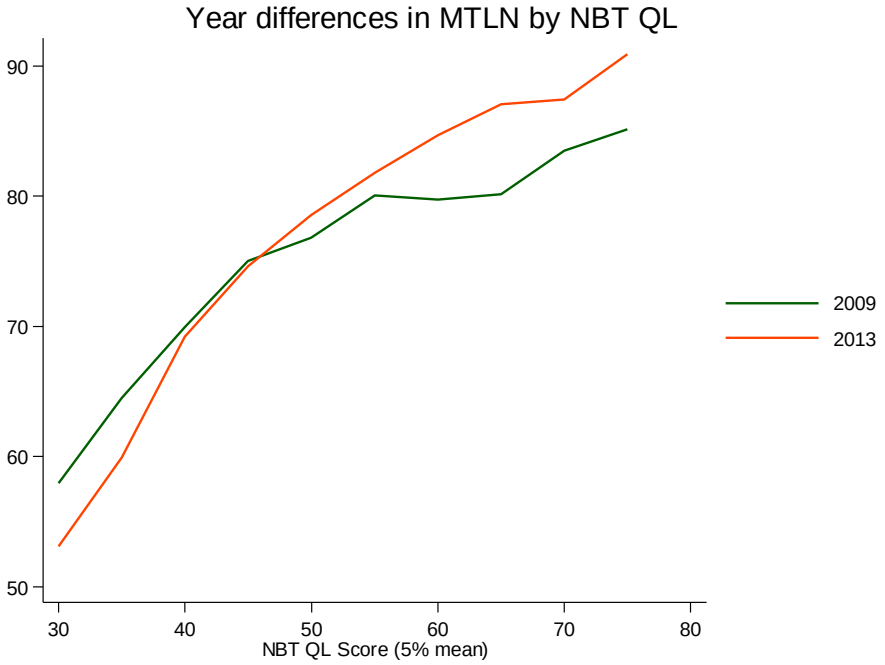



*Figure 15: Mean MTLN by mean NBT QL performance, 2009 and 2013*

## Five year trends by province

Provincial trends on mean NSC600 scores present a very interesting pattern. IEB candidates' performance is consistently at the top for most subjects while Eastern Cape hovers at the bottom in most subjects. This pattern is not new. However, it is worth noting that the gap between IEB candidates' performance and all other provinces in highest in NSC mathematics, followed by that in Physical Science. IEB candidates' performance in English language seems to be quite similar to that of the other provinces. It seems that NSC mathematics is a more discriminating subject in terms of identifying differences in ability than the other subjects.

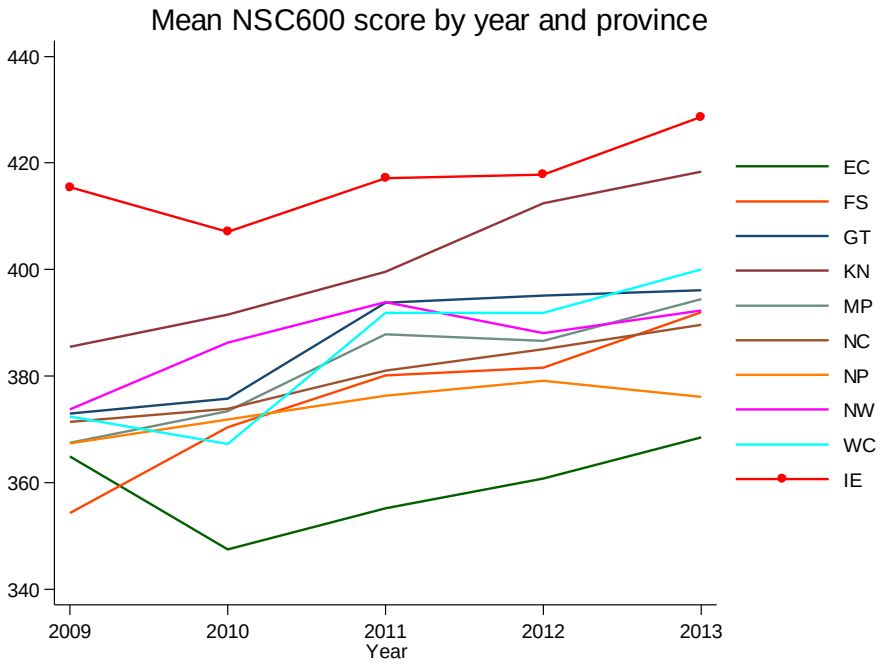



*Figure 16: Mean NSC600 score by year and province*

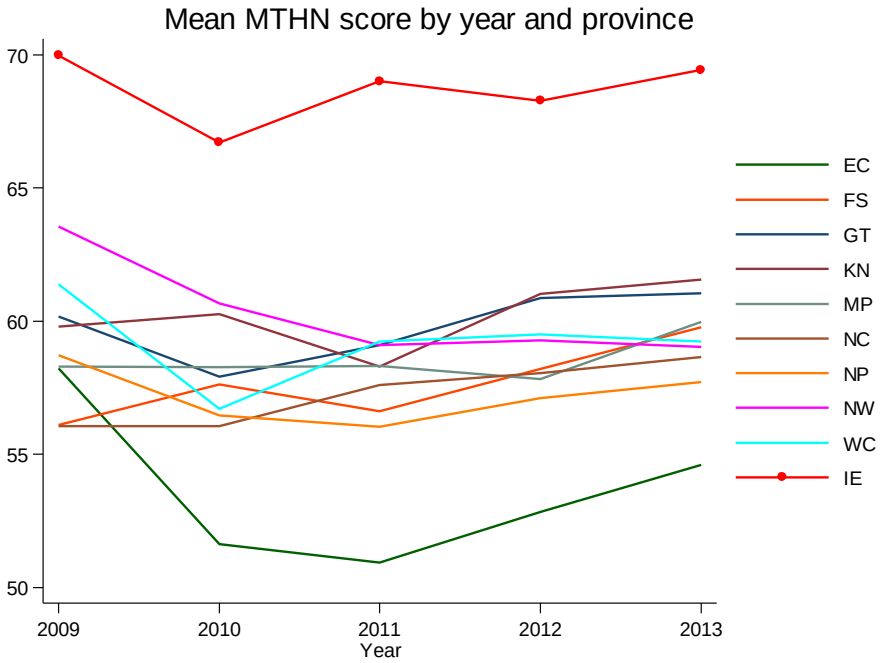


*Figure 17: Mean MTHN score by year and province*

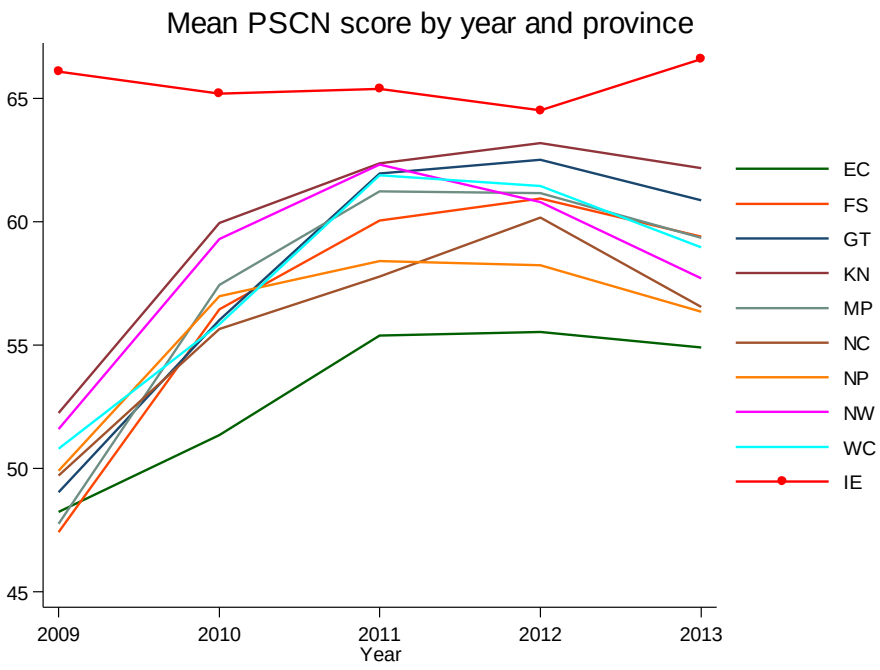



*Figure 18: Mean PSCN score by year and province*

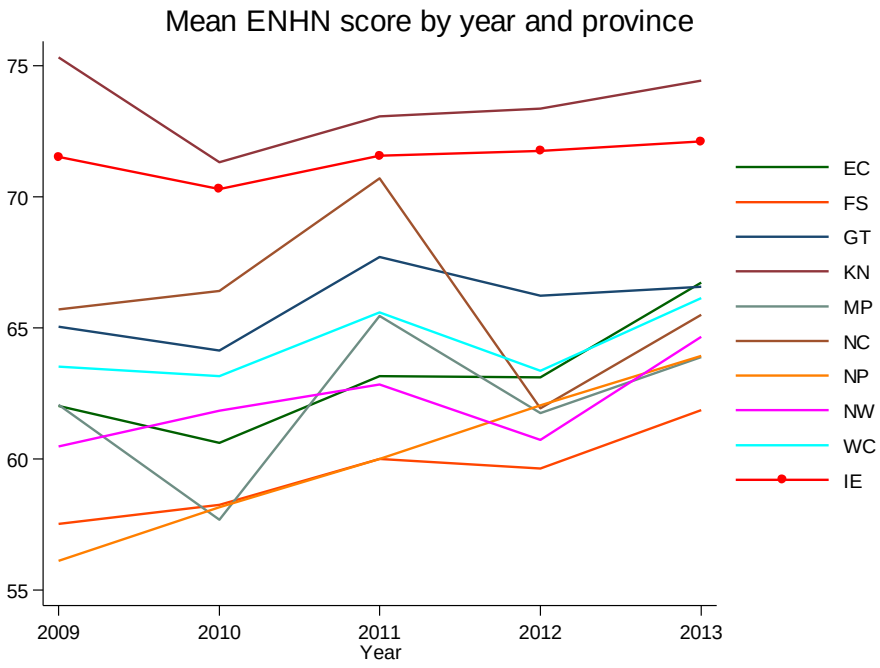
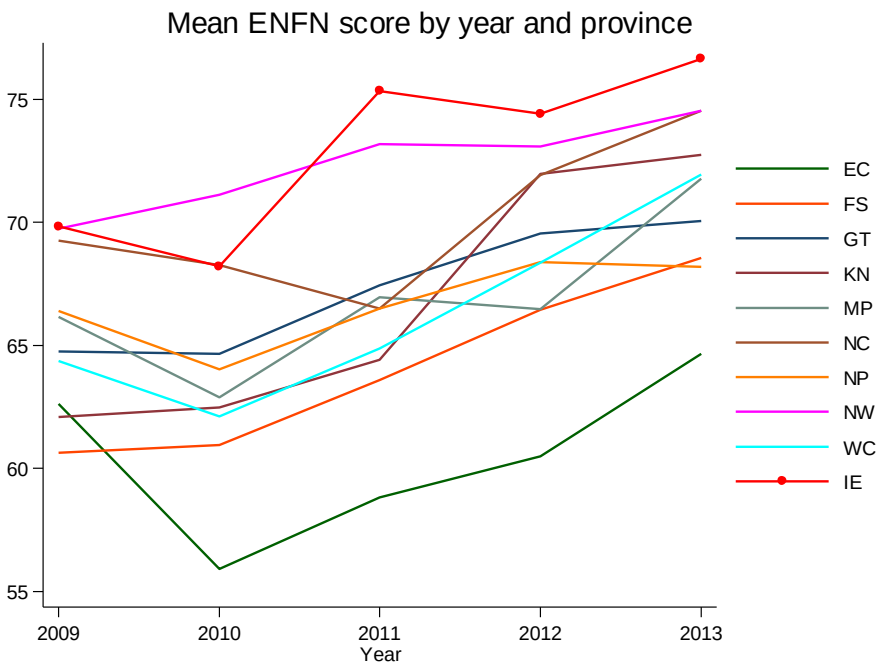



*Figure 19: Mean ENHN score by year and province*




The means NBT scores by province suggest a pattern that is slightly different from the NSC pattern and which needs some interrogation. IEB candidates dominate in terms of best performance as would be expected. However, the Western Cape Province is now second in all NBT scores. Although the difference from other provinces is a lot narrower than that shown by IEB candidates, this new difference is consistent. It may be worthwhile exploring the implications of these findings with the appropriate stakeholders.

*Figure 20: Mean ENFN score by year and province*

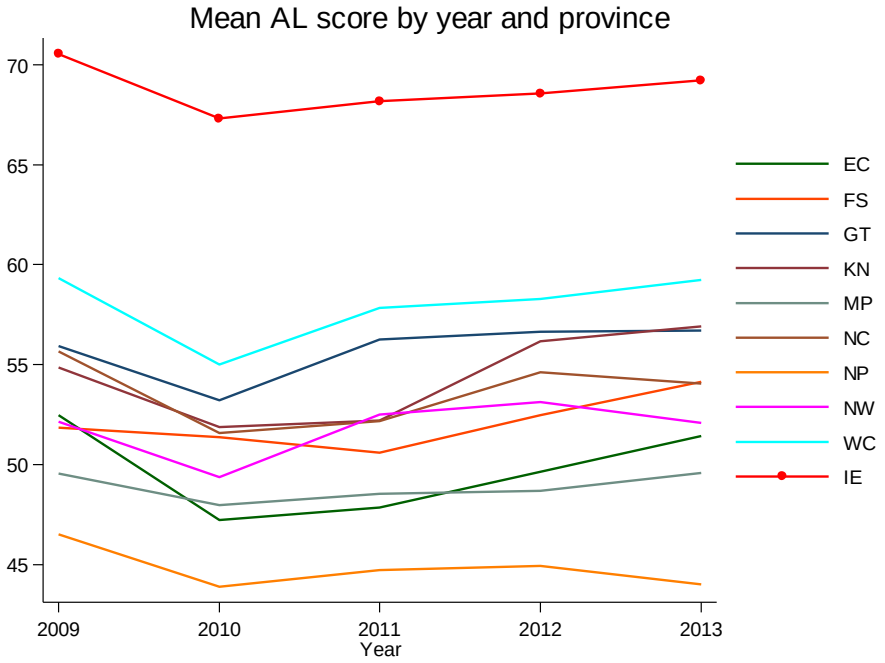



*Figure 21: Mean NBT AL score by year and province*

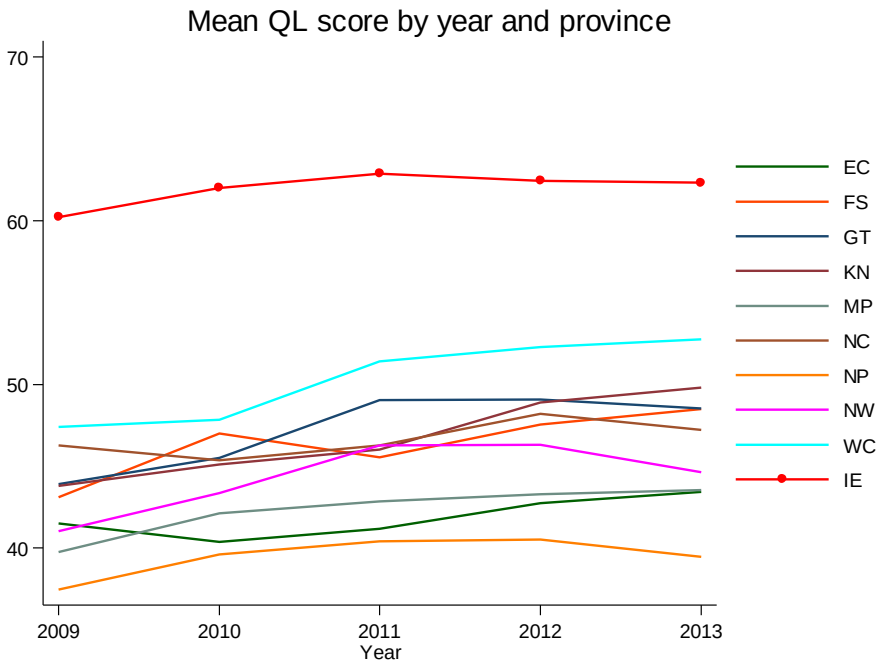



*Figure 22: Mean NBT QL score by year and province*

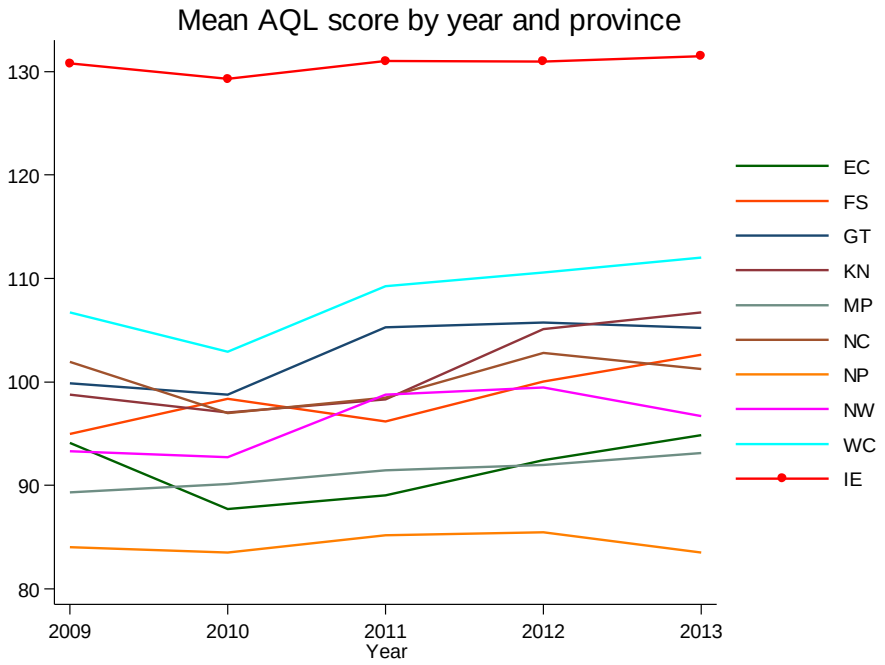



*Figure 23: Mean NBT AQL score by year and province*

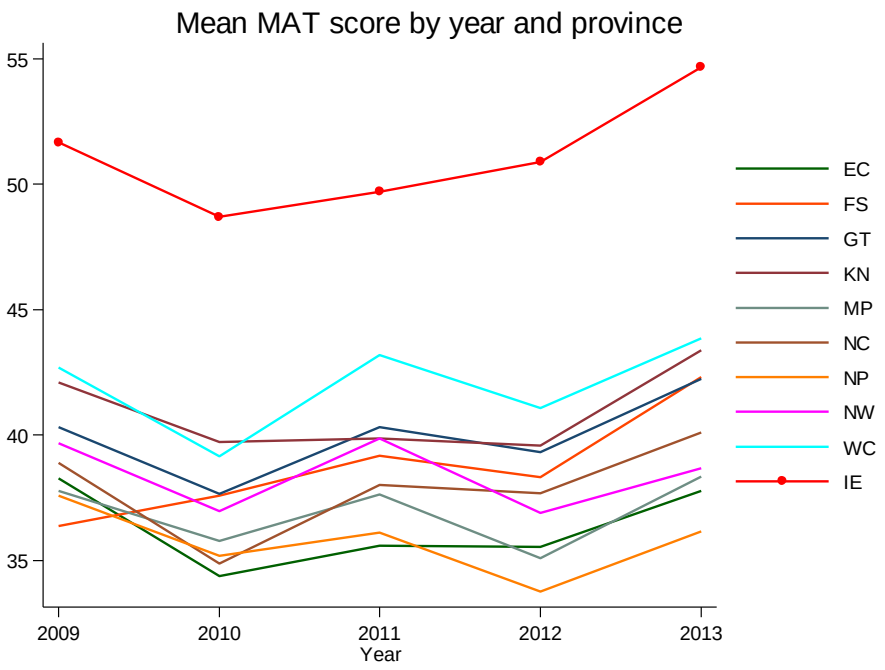



*Figure 24: Mean NBT MAT score by year and province*

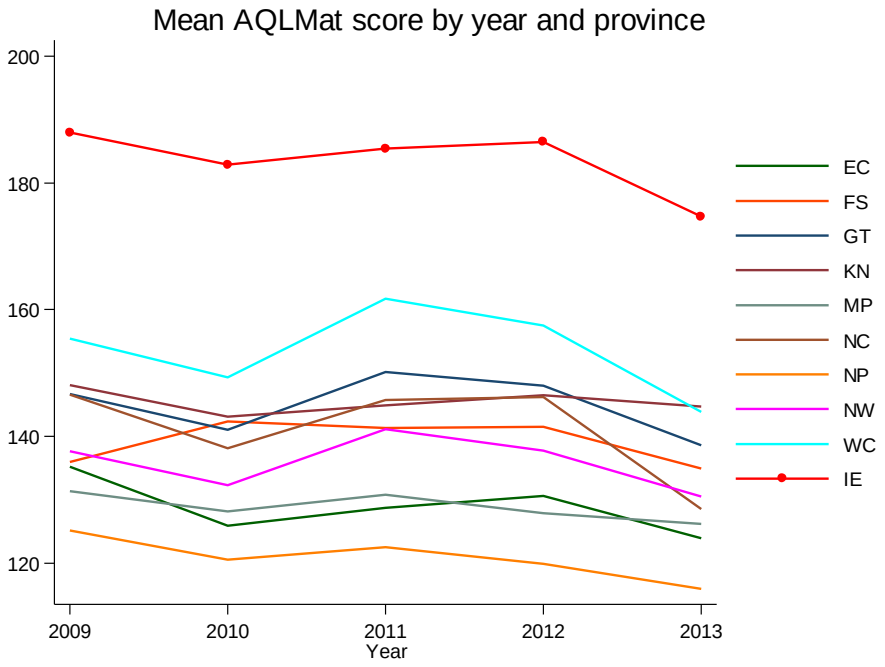



*Figure 25: Mean AQLMAT score by year and province*

## Testing differences in year means for NSC600, AQL and AQLMAT

T-test assesses whether the means of two groups are statistically different from each other in this case testing whether or not mean scores are the same for different years looking at NSC600, AQL and AQLMAT.

There are large differences in the year means for NSC600 ranging from -0.2 in 2009-10 to 22.8 in 2013. These year to year mean differences are all statistically significant at the .005 level, with the exception of the differences between 2009 and 2010. Such yearly large changes in mean scores clearly indicate grade inflation. Any other explanation would mean a sudden and unprecedented change in student ability for the better which is quite unlikely since university experience has recorded similar, and in some cases poorer, first year performance than previously.

The mean yearly differences in NBT AQL performance also increased during the same period from -2.9 to 4 in 2013. These differences are also statistically significant, except for the differences between 2012 and 2013. The magnitude of the t-test differences in mean yearly AQL is quite small compared to the NSC600 mean scores. This suggests some stability in AQL as a measure of candidates' ability and the results provide further evidence for NSC grade inflation.

The highlighted numbers are those indicating non-significant mean differences.

*Table 6: Testing differences in year means for NSC600 and NBT AQL*

| | 2009 | 2010 | 2011 | 2012 | 2013 |
|---|---|---|---|---|---|
| **Number with NSC600 and AQL** | 26 898 | 29 262 | 33 547 | 38 925 | 45 245 |
| **NSC600 mean** | 375.18 | 374.98 | 388.96 | 392.23 | 397.99 |
| **AQL mean** | 100.92 | 98.06 | 102.36 | 104.79 | 104.95 |

| **NSC600 differences in year means** | **2009** | **2010** | **2011** | **2012** | **2013** |
|---|---|---|---|---|---|
| **2009** | | -0.2 | 13.8 | 17.1 | 22.8 |
| **2010** | | | 14.0 | 17.3 | 23.0 |
| **2011** | | | | 3.3 | 9.0 |
| **2012** | | | | | 5.8 |
| **2013** | | | | | |

| **NSC600 ttest p-value matrix** | **2009** | **2010** | **2011** | **2012** | **2013** |
|---|---|---|---|---|---|
| **2009** | | 0.7294 | 0.0000 | 0.0000 | 0.0000 |
| **2010** | | | 0.0000 | 0.0000 | 0.0000 |
| **2011** | | | | 0.0000 | 0.0000 |
| **2012** | | | | | 0.0000 |
| **2013** | | | | | |

| **AQL differences in year means** | **2009** | **2010** | **2011** | **2012** | **2013** |
|---|---|---|---|---|---|
| **2009** | | -2.9 | 1.4 | 3.9 | 4.0 |
| **2010** | | | 4.3 | 6.7 | 6.9 |
| **2011** | | | | 2.4 | 2.6 |
| **2012** | | | | | 0.2 |
| **2013** | | | | | |

| AQL ttest p-value matrix | 2009 | 2010 | 2011 | 2012 | 2013 |
|---|---|---|---|---|---|
| 2009 | | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 2010 | | | 0.0000 | 0.0000 | 0.0000 |
| 2011 | | | | 0.0000 | 0.0000 |
| 2012 | | | | | 0.4195 |
| 2013 | | | | | |

*Table 7: Testing differences in year means for NSC600 and NBT AQLMAT*

| | 2009 | 2010 | 2011 | 2012 | 2013 |
|---|---|---|---|---|---|
| **Number with NSC600 and AQLMAT** | 15 325 | 20 003 | 23 998 | 30 887 | 36 415 |
| **NSC600 mean** | 387.69 | 386.43 | 398.9 | 399.21 | 405.02 |
| **AQLMAT mean** | 147.67 | 140.53 | 147.42 | 147.02 | 149.85 |

| NSC600 differences in year means | 2009 | 2010 | 2011 | 2012 | 2013 |
|---|---|---|---|---|---|
| 2009 | | -1.3 | 11.2 | 11.5 | 17.3 |
| 2010 | | | 12.5 | 12.8 | 18.6 |
| 2011 | | | | 0.3 | 6.1 |
| 2012 | | | | | 5.8 |
| 2013 | | | | | |

| NSC600 ttest p-value matrix | 2009 | 2010 | 2011 | 2012 | 2013 |
|---|---|---|---|---|---|
| 2009 | | 0.0841 | 0.0000 | 0.0000 | 0.0000 |
| 2010 | | | 0.0000 | 0.0000 | 0.0000 |
| 2011 | | | | 0.6067 | 0.0000 |
| 2012 | | | | | 0.0000 |
| 2013 | | | | | |

| AQLMAT differences in year means | 2009 | 2010 | 2011 | 2012 | 2013 |
|---|---|---|---|---|---|
| 2009 | | -7.1 | -0.3 | -0.6 | 2.2 |
| 2010 | | | 6.9 | 6.5 | 9.3 |
| 2011 | | | | -0.4 | 2.4 |
| 2012 | | | | | 2.8 |
| 2013 | | | | | |

| AQLMAT ttest p-value matrix | 2009 | 2010 | 2011 | 2012 | 2013 |
|---|---|---|---|---|---|
| 2009 | | 0.0000 | 0.5468 | 0.1014 | 0.0000 |
| 2010 | | | 0.0000 | 0.0000 | 0.0000 |
| 2011 | | | | 0.2480 | 0.0000 |
| 2012 | | | | | 0.0000 |
| 2013 | | | | | |

## Conclusion

The aim of this report was to examine the NSC scores and NBT scores for evidence of score (grade) inflation, using the NBT as the standard measure. The report began with a short review of the differences between standardized tests (like the NBTs) and the norm references tests (like the NSCs). This fundamental difference in measurement justifies the use of the NBTs as a standard in this report.

To facilitate easier comparison, the methodology used involved creating a composite NSC indicator (NSC600) and composite NBT scores AQL and AQLMAT which were then used in the analysis.

The results provided through graphs and tables clearly showed the following

1. NBT AQL is a stable measure. This is shown in the fairly consistent correlations between NBT AQL and NSC subjects.
2. Year to year increases in NSC 600 performance seem quite large, yet these are not supported by NBTAQL performance scores. This strongly suggest NSC score (grade) inflation.
3. Candidates presenting ENFN are scoring approximately 10 percentage points higher than 'AL equivalent' ENHN candidates.
4. Further statistical tests, comparing the yearly differences in means for NSC scores and NBT AQL using t-tests indicated that increases in NSC 600 have been very large compared to those for NBT AQL which have increased at a moderate rate. These findings also suggest that the yearly increase in NSC scores is too large to be accounted for by an increase in student ability. Rather this increase can only be explained by grade inflation, in other words, performance on the NSCs are inflating grades to make candidates reflect better performance than suggested by their actual ability.

   The Implications of this report is that there is need for further work on score (grade) inflation in the NSC using the NBT as a standard. If score (grade) inflation is confirmed, then there is need to review the NSC examinations, marking procedures and grading to ensure that the NSCs can better reflect a candidates' ability.

   Notes:
   1. The NSC is a high stakes test – behaviour of students and teachers are affected.
   2. The Annual National Assessments (ANA's) and the NBTs seem to provide similar information about the abilities of learners. While performance on the NSC indicate that learners' abilities and preparedness are different to the information provided by either the NBT or for that matter the ANAs. This has serious consequences for using the NSC for HE admissions especially in making placement decisions.
   3. Making inferences about the larger domain of Mathematics based on performance in the NSC Mathematics (MTHN) is hugely problematic –being prepared for the MTHN examination and doing well on it does not necessarily mean that students have a high ability in mathematics or that they are prepared for the mathematics demands of higher education.
   4. Score inflation, teacher behaviour and alignment.
      a. Reallocation – how time is used. Time is used to improve performance on the examination rather than on improving ability in the domain of interest.
      b. Alignment
      c. Coaching
   5. High stakes decisions, such as higher education admissions, is best made on as much information as is available rather than being based on a single examination score