



NATIONAL BENCHMARK TEST AS A BENCHMARK TOOL

Radley Mahlobo

Vaal University of Technology

South Africa

radley@vut.ac.za

Abstract - This paper reacts out of concern for the high discrepancy between the 2014 National Benchmark Test (NBT) and the 2013 National Senior Certificate (NSC) examination results. Since both the NBT and NSC results impact on the ultimate placement of students in tertiary education, the paper focuses on NBT mathematics test (MAT) in its current format as a benchmark tool for establishing mathematical skills of prospective tertiary students. It also undertakes a comparative analysis of both assessment modes in terms of predicting performance in first-year tertiary mathematics. Intensive literature survey seems to discredit the multiple choice format of MAT, and consequently MAT as a valid benchmarking tool. Analysis of the 2014 MAT results and those of first-year examination seems to question the status of MAT as a predictive benchmark tool for prospective tertiary students.

Keywords: National Benchmark Test, National Senior Certificate & Mathematics

1.1. Introduction

The National Benchmark Tests (NBTs) were first introduced in 2005 by Higher Education South Africa (HESA). These tests, according to HESA, had 3 purposes:

- To assess entry-level academic literacy and mathematics skills.
- To assess the relationship between entry-level skills and school-level exit results.
- To provide institutions with additional information in the admission and placement of entry-level students.

Two tests, namely Academic and Quantitative Literacy Test (AQL) and Mathematics Test (MAT) were taken by prospective tertiary students. The AQL is multiple-choice and has a total of 3 hours of writing time. This particular test, a combination of academic literacy and quantitative literacy, needs to be taken by all applicants regardless of what they plan to study. MAT is written by applicants to programmes for which maths is a requirement. It is also multiple-choice and the student has 3 hours to complete it.

Concern has been expressed at the discrepancy between the National Senior Certificate (NSC) exams, which most students passed, and the National Benchmark Tests (NBTs), which the majority did not pass. The NSC is run by the National Department of Basic Education (DoBE), while NBT is run from University of Cape Town (UCT). Prospective tertiary students register with UCT in order to write NBTs.

Stakeholders questioned what the consequences were to failing the Benchmark Tests, whether these were used to indicate what level of support was required, or whether they were used to “gate-keep” and block some students from admission. The response, among others, was that the NBTs had to tell universities to institutionalise their support mechanisms; they were not supposed to be used to turn students away.

Stakeholders commented that there needed to be greater synergy between the NSC level and the NBT level, so that there could ideally be one test for entrance, qualification and support assessment. Comments on the 2008 exam papers were received from chief markers, universities, teachers unions and the general public, and generally agreed that most NSC question papers were of an appropriate standard, based on different cognitive levels, catered for low, medium and high achievers, and adequately assessed the content as stipulated in the National Senior Certificate.

1.2. Observations leading to research questions.

The NBT mathematics test (MAT), as a benchmark test, sets a national standard against which the mathematical skills, among others, of an entry-level tertiary student, are measured. The National Senior Certificate (NSC) exams, by being used for tertiary admissions, are automatic tools not only to establish the mathematical skills of a prospective tertiary student, but also to influence placement of the students in different mathematically-orientated tertiary study fields. Both forms of assessment seem to have the same target – establishment of mathematical skills of prospective tertiary students. “For many good reasons the MATs do not attempt to replicate the NSC mathematics exam papers. Clearly the NSC mathematics exams and the MATs should be seen as complementary forms of assessment” (NBT, 2013). The concern of this paper is that a high score discrepancy between two sets of assessment that are intended to complement each other deserves attention. This paper focuses on the NBT as a complementary tool for decision making within the framework of the expected skills of a citizen envisaged by the South African education system. The policy framework describes the anticipated skills of the citizen, and describes Critical Outcomes the citizen must display as a result of acquisition of the skills. Specifically, this paper asks two questions in relation to assessment mathematical skills:

1.3. Research questions

- 1.3.1. To what extent is MAT in its current form a suitable tool to assess envisaged academic mathematics skills?
- 1.3.2. To what extent are the MAT results better predictors of the student performance in their first-year mathematics studies as opposed to NSC results?

The second question arises from the assumption that assessment of academic mathematical skills has a predictive value of student performance in academic mathematics. This, in fact, is expressed in the NBT document: “MATs elect to focus on those aspects of the school curriculum that have a greater bearing on performance in first-year mathematics” (NBT, 2013:2).

Since both NSC and MAT are embedded in the national curriculum, we look at the national context of the envisaged educational citizen.

2. Literature survey

2.1. South African context of envisaged skills.

Education and Training in South Africa has seven Critical Outcomes (COs), which derive from the Constitution (DoE, 2008:10). Each of the seven critical outcomes describes an essential characteristic of the type of South African citizen the education sector hopes to produce.

In South Africa, “Curriculum and Assessment Policy Statements” (CAPs) refer to the policy documents stipulating the aim, scope, content and assessment for each subject listed in the *National Curriculum Statement Grades R – 12* (NCS). According to CAPs (2011:6), the NCS aims to produce learners that are able to fulfill the following seven COs:

- identify and solve problems and make decisions using critical and creative thinking;



- work effectively as individuals and with others as members of a team;
- organise and manage themselves and their activities responsibly and effectively;
- collect, analyse, organise and critically evaluate information;
- communicate effectively using visual, symbolic and/or language skills in various modes;
- use science and technology effectively and critically showing responsibility towards the environment and the health of others; and
- demonstrate an understanding of the world as a set of related systems by recognising that problem solving contexts do not exist in isolation.

There are many angles from which to look at mathematical skills linked to achievements of the COs. For instance, the assessment of the second CO will take cognisance of the fact that “in the world of work most problem solving involves teamwork, and while a mathematician will bring special skills to a team, he will also have to function as a member of that team and to be aware of the issues involved in that” (Challis et al 2002). Assessment of the second CO may have a different focus as compared to those of the other COs. In this paper focus is on the first critical outcome: identify and solve problems and make decisions using critical and creative thinking. Attention of this paper will be placed on the mathematical “critical thinking” aspect of the Critical Outcome. Before proceeding, we describe critical thinking.

2.2. Critical Thinking

Despite the widespread attention on critical thinking, no clear-cut definition has been identified (Ou Lydia *et al*, 2014). Among the definitions of critical thinking from current frameworks of learning outcomes is that by Binkley *et al* (2012). They define critical thinking as ways of thinking that incorporate problem solving and decision making. They claim that critical thinking can be categorised into knowledge, skills and attitudes/values/ethics. Knowledge includes a) reason effectively, use systems thinking, and evaluate evidence; b) solve problems and c) clearly articulate. Skills include a) reason effectively and b) use systems thinking. Attitudes/values/ethics include a) make reasoned judgments and decisions, b) solve problems, and attitudinal disposition. Halpern (2003:6) defined critical thinking as “the use of those cognitive skills or strategies that increase the probability of a desired outcome. It is used to describe thinking that is purposeful, reasoned, and goal directed – the kind of thinking involved in problem solving, formulating inferences, calculating likelihoods, and making decisions, when the thinker is using skills that are thoughtful and effective for the particular context and type of thinking tasks”.

According to Ou Lydia *et al* (2014), critical thinking is one of the most important skills deemed necessary for college graduates to become effective contributors in the global workforce. Critical thinking is one of the most frequently discussed higher-order skills, believed to play a central role in logical thinking, decision making, and problem solving (Butler, 2012;Halpern, 2003). It is also a highly contentious skill in that researchers debate about its definition; its amenability to assessment and the evidence of its practical impact on people’s academic achievement.

In a survey conducted by the Association of American Colleges and Universities (AAC & U, 2011), 95% of the chief academic officers from 433 institutions rate critical thinking as one of the most important intellectual skills for their students. 81% of the employees AAC & U (2011) wanted

colleges to place stronger emphasis on critical thinking. In a study conducted by Educational Testing Service (ETS, 2013), vice presidents of academic affairs from more than 200 institutions were interviewed regarding the most commonly measured general educational skills, and critical thinking was one of the most frequently mentioned competencies considered essential for both academic and career success.

Despite the widespread attention on critical thinking, no clear-cut definition has been identified (Ou Lydia *et al*, 2014). Among the definitions of critical thinking from current frameworks of learning outcomes is that by Binkley *et al* (2012). They define critical thinking as ways of thinking that incorporate problem solving and decision making. They claim that critical thinking can be categorised into knowledge, skills and attitudes/values/ethics. Knowledge includes a) reason effectively, use systems thinking, and evaluate evidence; b) solve problems and c) clearly articulate. Skills include a) reason effectively and b) use systems thinking. Attitudes/values/ethics include a) make reasoned judgments and decisions, b) solve problems, and attitudinal disposition. Halpern (2003:6) defined critical thinking as “the use of those cognitive skills or strategies that increase the probability of a desired outcome. It is used to describe thinking that is purposeful, reasoned, and goal directed – the kind of thinking involved in problem solving, formulating inferences, calculating likelihoods, and making decisions, when the thinker is using skills that are thoughtful and effective for the particular context and type of thinking tasks”.

3. Theoretical Framework

While acknowledging that there is no clear-cut definition of critical thinking, this paper notes that critical thinking can be determined through assessment of skills used in problem solving and decision making (Binkey *et al*, 2012, Butler, 2012; Halpern, 2003). This notion is consistent with some of the objectives of MAT: “The MATs require writers to demonstrate sufficient understanding of concepts to enable them to apply those concepts in a variety of context” (NBT, 2013:2). “The MAT focuses more on the knowledge and skills taught at school level, but are explicitly designed to measure the preparedness of candidates for higher education. These higher-order skills underlie success in Mathematics in Higher Education” (NBT, 2013:2).

It is the contention of this paper that the ‘problem solving – decision making’ notion of critical thinking can be demonstrated by the achievement of the Critical Outcome: “learners must identify and solve problems and make decisions using critical and creative thinking”.

In the following section we briefly discuss how this investigation will be conducted.

4. Research methodology

4.1. Research design

This research paper has two points of focus. The first focus, responding to the first research question, will be on scrutinising the MAT with a view of establishing its suitability as a tool to determine academic mathematical skills of prospective tertiary students. Literature will be reviewed to respond to the first question, and results will follow from the analysis of the literature.

The approach needed to answer the second research question will be a quantitative approach. 2013 MAT and NSC exam results will be analysed. Means and levels of significance will be used to compare 2013 MAT results with 2013 First-year examination result on one hand, and 2013 NSC examination results with 2013 First-year examination result on the other hand, in order to establish which of the two results will better predict performance of the students in first-year mathematics.

4.2. Population and samples for second question

The answer to the second question will involve 184 prospective tertiary students, belonging to a University of Technology in South Africa, who wrote 2014 MATs, 2013 NSC examinations and first-year mathematics examination.

4.3. Data collection instrument for second question.

MATs and NSC examinations were the instruments used to collect data.

4.3.1. Validity of the instrument:

Validity refers to how well a test measures what it is purported to measure (Colin & Wren, 2005). The tests (MAT and NSC) all adhere to the same set of specifications, and are as equivalent as possible before they are written (NBT, 2013:3). The contention of this paper is on whether MAT as a multiple choice assessment is a valid form of assessment. The tests were thus content validated.

4.3.2. Reliability of the instrument

Reliability is the degree to which an assessment tool produces stable and consistent results (Colin & Wren, 2005). For the reliability of MAT and NSC examination K-R 21 formula was used and yielded acceptable reliability coefficients. It was thus concluded that the instruments were reliable.

4. 4. Data Analysis for second question

Quantitative data analysis will be used. As mentioned above, means and levels of significance will be used to compare 2013 MAT results with 2013 First-year examination result on one hand, and 2013 NSC examination results with 2013 First-year examination result on the other hand.

Before focusing on the results, we compare the NBTs with the NSC examinations, in order to be able to highlight MAT as a decision making tool.

5. A closer look at MATs and NSC mathematics examinations.

5.1. Contrasting the two

While it is accepted that both MAT and NSC exam are equivalent, it is also helpful to contrast them. Knowing that many applicants to universities need to write the NBTs as early as the end of May, the MATs exclude topics that are unlikely to have been taught by that time.

The questions in the MAT are embedded in the concepts set out in the CAPS statement, but the tests are not constrained to testing everything covered by the CAPS. The NSC exams are written by all grade 12s, while the MATs are written only by prospective students who intend studying courses for which mathematics is a requirement.

While the MATs cannot test anything outside the school curriculum, they are not constraint to include all school mathematics topics, and thus elect to focus on those aspects of the school curriculum that have greater bearing on performance in first-year mathematics courses (NBT, 2013:2).



The CAPS FET Band Mathematics Grades 10 – 12 document specifies a taxonomy of categories of mathematical demand:

- Knowing (recall or basic factual knowledge) about 20%.
- Performing routine procedures: about 35%.
- Performing complex procedures: about 30%.
- Problem solving: about 15% (NBT, 2013:2)

The MATs are also differentiated in terms cognitive levels, starting with lower-order questions in order to facilitate an easy introduction into the test, and then progressing to questions that have greater cognitive demand.

- The highest level (counting for about 8%) comprises questions that involve greater insight.
- The lowest level (about 45%) comprises questions that involve knowledge, recall and application of simple procedures (NBT, 2013:3).

5.2. Differences between the NSC mathematics examination and the MAT.

The MATs do not cue the writer in any way. For example, in an NSC paper the following might appear: Given a sketch, learners are asked: “Calculate the gradient of AC. Hence determine the equation of BN (where BN is shown on the sketch to be perpendicular to AC)”. In the MAT the sketch would also be presented, but would then be followed by: “The equation of BN is... with four options to choose from” (NBT, 2013:3).

No indication is given as to whether a question should be dealt with using geometrical or algebraic reasoning. The fact that mathematics often requires learners to integrate many different skills and concepts in any given problem means that individual questions will assess across a range of mathematical competencies. For example, a question dealing with the graphical representation of a function may also assess spatial and algebraic competencies. This means that writers must have a deep understanding of mathematics and know what reasoning is appropriate in a given context, they will need these skills in Higher Education.

“It may be assumed that multiple choice testing does not allow writers to obtain part marks for their reasoning in cases where they have reasoned correctly until the last step and then made a final careless mistake. This criticism is understandable, but the NBTP review process has, over a period of more than five years, made it possible to fine-tune the process of creating options for which this is unlikely. Firstly, if numerical reasoning is involved, the numbers are simple (sufficiently so to make to make calculators unnecessary). Secondly, the options given provide one correct answer and others which are unlikely to have been reached by making careless mistakes” (NBT, 2013:3).

What results were obtained by this investigation?

6. Results:

6.1. Research question: Is MAT in its current form a suitable tool to assess envisaged academic mathematics skills?

6.1.1. MAT competency skills versus skills of envisaged citizen

To determine whether learners are able to make transition between mathematics at secondary and tertiary levels, certain competencies were identified by the MAT for assessment (NBT, 2013:5). It seems the formulation of MAT achievement levels (Basic, Intermediate and Proficient) is consistent with the requirements of the South African educational policy document. Skills associated with achievement of these levels are captured by the critical outcome: identify and solve problems and make decisions using critical thinking. The following table describes the achievement levels (NBT, 2013):

Basic Competency Skill	Intermediate Skill	Proficient
Test writers performing at this level will be able to: <ul style="list-style-type: none"> • Apply simple concepts • Use known procedures in familiar situations 	Test writers performing at this level will be able to perform at the Basic level and in addition to be able to: <ul style="list-style-type: none"> • Select strategies to solve problems. • Integrate skills, concepts and procedures. 	Test writers performing at this level will be able to perform at the Intermediate level and in addition to be able to: <ul style="list-style-type: none"> • Demonstrate in-depth knowledge of the relevant mathematical concepts and competence in multi-step procedures represented in the framework.

The achievement levels associated with each of the mathematics topics were also specified. For instance the following table indicates the achievement levels for the topic 'Trigonometry'.

Trigonometry		
Basic Competency Skill	Intermediate Skill	Proficient
Test writers should be able to: <ul style="list-style-type: none"> • Define specified trigonometric ratios in right-angled trigonometry • Recognise trigonometric graphs. • Perform basic calculations with trigonometric ratios. • Solve simple trigonometric equations. • Recognise and use identities, compound-angle formulae and reduction formulae 	Test writers should be able to: <ul style="list-style-type: none"> • Apply trigonometric ratios to solve two-dimensional problems. • Understand the properties of trigonometric graphs, including translations of these graphs. • Apply the sine, cosine and area rules in simple contexts. • Solve non-routine trigonometric equations. • Apply trigonometric equations, reduction formulae and knowledge of special angles to solve 	Test writers should be able to: <ul style="list-style-type: none"> • Interpret graphs of trigonometric functions, individually and in relation to one another. • Apply trigonometric concepts to solve non-routine problems in two- and three-dimensional contexts.

	problems.	
--	-----------	--

If the above achievement levels are taken as checklists for assessment of mathematical skills, then they can go a long way to achieving the stated goals of helping with placement. In other words, assessment of the stated achievement levels will make the MAT a very useful tool to assess the mathematical skills of the prospective tertiary students.

6.1.2. MAT as a multiple choice test

Multiple choice items consist of a question or incomplete statement (called a stem) followed by 3 to 5 response options. The correct response is called the key while the incorrect response options are called distracters. Multiple-choice testing became popular in the 1900's because of the efficiency that it provided (Swartz, 2006). As the influence of psychometricians grew in 1926, the psychometric test became a multiple-choice test (Matzen and Hoyt, 2006).

Advantages of multiple-choice tests include how quickly tests can be graded compared to others. It is much more cost effective than having to read over written answers which take time and possibly training depending on who is employed to grade them (Holtzman, 2008). The large number of questions makes it possible to test a broad range of content and provides a good sample of the test taker's knowledge, reducing the effect of "the luck of the draw" (Livingstone, 2009).

Questions that require the test taker to produce the answer, rather than simply choosing it from a list, are referred to as constructed-response questions. Although constructed-response items have great face validity and have the potential to offer authentic contexts in assessment, they tend to have lower levels of reliability than multiple-choice items for the same amount of time (Lee, Liu, & Linn, 2011). Studies show high correlations of multiple-choice items and constructed-response items of the same constructs (Klein et al., 2009). Given that there may be situations where constructed-response items are more expensive to score and that multiple-choice items can measure the same constructs equally well in some cases, one might argue that it makes more sense to use all multiple-choice items and disregard construct-response items.

However, there is a down side to assessment using multiple choice questions. One is limited feedback from student response to correct errors in the student understanding. Consider the following solutions of two students:

Question : Evaluate $\frac{dy}{dx}$ at $x = \frac{\pi}{4}$, given that $y = (\cos x + \sin x)^2$

Student 1 solution

$$y = (\sin x + \cos x)^2 \Rightarrow y = \sin^2 x + 2 \sin x \cos x + \cos^2 x = 1 + 2 \sin x \cos x$$

$$\therefore \frac{dy}{dx} = \frac{d}{dx} (1 + 2 \sin x \cos x) = [2 \cos^2 x - 2 \sin^2 x] \therefore \left. \frac{dy}{dx} \right|_{x=\frac{\pi}{4}} = 2 \cos^2 \frac{\pi}{4} - 2 \sin^2 \frac{\pi}{4} = 0$$

Student 2 solution

$$y = (\sin x + \cos x)^2 \Rightarrow y = \sin^2 x + \cos^2 x = 1$$

$$\therefore \frac{dy}{dx} = \frac{d}{dx} (1) = 0 \therefore \left. \frac{dy}{dx} \right|_{x=\frac{\pi}{4}} = 0$$

Student 2 solution is based on a serious algebraic misconception $(\sin x + \cos x)^2 = \sin^2 x + \cos^2 x$ but it leads to the correct answer 0. The marker will not be able to pick up the misconception. In this case the mathematical skills of the writer cannot be established by the multiple choice questions.

Many skills that schools teach are too complex to be measured effectively with multiple-choice questions. A multiple-choice test for mathematics students can determine whether they can solve many kinds of questions, but it cannot determine whether they can construct a mathematical proof (Livingstone, 2009). Unfortunately, the whole focus of this paper is on mathematics as a school subject. As a practising mathematics lecturer at a tertiary institution, I have observed with disappointment a student culture in our tertiary institutions. Students are not interested in understanding how certain formulae are derived. They are more interested in using the formula to solve routine problems. Unless they are given the formula, they are unable to solve the problems. This usually results in poor transfer of mathematical skills to other contexts.

Students who cannot state the general scientific principle illustrated by a specific process in nature may have no trouble recognising that principle when they see it stated along with three or four others (Livingstone, 2009).

In academic subjects, there is usually a strong tendency for the students who are stronger in the skills measured by multiple-choice questions to be stronger in the skills measured by constructed-response test questions. But if all the students improve in the skills tested by constructed-response test questions, their performance on the multiple-choice questions may not reflect that improvement (Livingstone, 2009). The above example applies here. If student 2 can rectify the misconception that $(\cos x + \sin x)^2 = \cos^2 x + \sin^2 x$ and ultimately get the correct response as student 1, multiple choice assessment will not pick the self-correction. Consequently the improvement associated with the self-correction will not be established by the assessment. In other words, the use of multiple choice questions will make it difficult to establish the mathematical skills of the prospective students.

When multiple-choice tests are used as the basis for important decisions, teachers have a strong incentive to emphasise the skills tested by the questions on those tests. An advice given to NBT teachers: "It might be helpful to give learners some guidelines regarding how to deal multiple choice tests" (NBT, 2013:7), will probably perpetuate the incentive. With a limited amount of class time available, they have to give a lower priority to the kinds of skills that would be tested by constructed-response test questions (Livingstone, 2009). If the skills tested by the multiple choice questions are found to be inadequate, then observations like "In engineering mathematics modules, where a frequent complaint is that students can do the maths in Maths but not in Engineering" (Challis et al, 2002) will be experienced.

With construct-response, it is possible to create more authentic contexts and assess students' ability to generate rather than select responses. In real-life situations where critical thinking skills need to be exercised, there will not be choices provided. In the case of critical thinking, construct-response items could be a better proxy of real-life scenarios than multiple-choice items (Livingstone, 2009).

One of the greatest problems in constructed-response testing is the time and expense involved in scoring. In recent years, researchers have made a great deal of progress in using computers to score the responses. Four scoring engines – computer programmes for automated scoring – have been developed in the past few years. They are called e-rater (for "essay rater") (Attal & Burstein, 2006), c-rater (for "content rater") (Leacock & Chodorov, 2003), and m-rater (for math rater"). The m-rater engine is a scoring engine for responses that consist of an algebraic expression (e.g. a formula), a plotted line or a curve on a graph, or a geometric figure. To score an algebraic expression, the m-rater engine determines whether the formula written by the test taker is algebraically equivalent to



the correct answer. The m-rater engine scores a straight-line graph by transforming the line into an algebraic expression; it scores a curved-line graph by testing the curve for correctness at several points (Livingstone, 2009).

To NBT credit, there are precautionary measures specified in MAT. For instance, “Questions in which it is possible to eliminate options by substitution are deliberately avoided. So, for example, there will not be questions asking for a specific solution to an equation, because it is easy to substitute each of the given options and find the correct answer by elimination” (NBT 2013:8). NBT also mentioned: “It may be assumed that multiple choice testing does not allow writers to obtain part marks for their reasoning in cases where they have reasoned correctly until the last step and the made a final careless mistake. This criticism is understandable, but the NBTP review process has, over a period of more than five years, made it possible to fine-tune the process of creating options for which this is unlikely. Firstly, if numerical reasoning is involved, the numbers are simple. Secondly, the options given provide one correct answer and three others which are unlikely to have been reached by making careless mistakes” (NBT, 2013:3). However, these precautionary measures are not sufficient to eliminate the inadequacy of the MAT as a multiple choice assessment of the academic mathematical skill. The fact that MAT is written before the completion of the syllabus does not help, either.

6.2. Question 2: Are the MAT results better predictors of the student performance in their first-year mathematics studies as opposed to NSC results?

6.2.1. Introduction

The predictive value of MAT can be inferred from the statement that: “The MAT elects to focus on those aspects of the school curriculum that have a greater bearing on performance in first-year mathematics courses” (NBT, 2013:2). It has already been mentioned that concern has been expressed at the discrepancy between the National Senior Certificate (NSC) exams, which most students passed, and the National Benchmark Tests (NBTs), which the majority did not pass. This could imply, among others, that the NBT standard is indicative of tertiary performance while the NSC standard is low, or that the NSC standard is indicative of tertiary performance and NBT is too high. That has necessitated a comparative look at the predictive value of the two assessments in relation to first-year mathematics performance.

We have so far dealt with the research design and methodology that culminated in the production of the data needed to facilitate the answering of the research questions. In this chapter, the quantitative data is represented and interpreted. The t-test was used to compare the means two pairs of tests, namely NSC (Gr 12) and first-year engineering mathematics (1E) on one side and MAT (NBT) and first-year engineering mathematics (1E) on the other side. The difference, if any, of the means of the compared tests was considered to be significant if the *significance level p* was at most 0.05 ($p \leq 0.05$). The focus will now be placed on representation and interpretation of and reflection on the data. Before we consider Gr 12 versus 1E and NBT versus 1E, we confirm the discrepancy between NBT and Gr 12:

6.2.2. a) NBT versus Gr 12

t-Test: Paired Two Sample for Means

	NBT	Gr 12
Mean	34.8913	66.25



Variance	92.23949	143.2377049
Observations	184	184
Pearson Correlation	0.29784	
Hypothesized Mean Difference	0	
df	183	
t Stat	-32.9154	
P(T<=t) one-tail	4.27E-79	
t Critical one-tail	1.653223	
P(T<=t) two-tail	8.55E-79	
t Critical two-tail	1.973012	

The discrepancy between the mean of NBT (34, 89) and that of Gr 12 (66, 25) is of statistical significance ($p = 8,55 \times 10^{-79} < 0,05$). This confirms the high discrepancy between the two modes of assessment.

6.2.3. Comparison results: 2013 NSC mathematics (Gr12) ($N = 184$) versus first-year (1E)2014 ($N = 184$) and MAT (NBT) ($N = 184$) versus first-year(1E) ($N = 184$)

t-Test: Paired Two Sample for Means

	<i>Gr 12</i>	<i>1E</i>
Mean	66.25	50.85869565
Variance	143.2377	311.6083393
Observations	184	184
Pearson Correlation	0.35597	
Hypothesized Mean Difference	0	
df	183	
t Stat	11.96565	
P(T<=t) one-tail	4.77E-25	
t Critical one-tail	1.653223	
P(T<=t) two-tail	9.53E-25	
t Critical two-tail	1.973012	

t-Test: Paired Two Sample for Means

	<i>NBT</i>	<i>1E</i>
Mean	34.8913	50.8587
Variance	92.23949	311.6083
Observations	184	184
Pearson Correlation	0.26305	
Hypothesized Mean Difference	0	
df	183	
t Stat	-12.2103	
P(T<=t) one-tail	9.11E-26	
t Critical one-tail	1.653223	



P(T<=t) two-tail	1.82E-25
t Critical two-tail	1.973012

What is emerging from the data is that both MAT results and those of NSC examinations are not good predictors of the first-year examination results. Average of NSC examination results (66, 25) is significantly ($p = 9,53 \times 10^{-25} < 0,05$) higher than that of first-year examination result. Average of MAT (34, 89), on the other hand, is significantly ($p = 1,82 \times 10^{-25} < 0,05$) lower than that of the first-year examination results.

7. Discussion and recommendations.

7.1. Is MAT in its current form a suitable tool to assess envisaged academic mathematics skills?

Achievement levels as specified in the MAT were found to be suitable to potentially extract the desired information to establish the mathematical skills of the prospective students. The sample questions were also found to be of good standard. However, as a multiple choice form of assessment, MAT has been found to be inadequate as a tool to assess mathematical skills of prospective tertiary students.

It is recommended that MAT be changed from its current multiple choice format to a constructed-response format.

7.2. Question 2: Are the MAT results better predictors of the student performance in their first-year mathematics studies as opposed to NSC results?

The issue of the discrepancy between MAT and NSC results raised a number of questions that all relate to whether MAT can be considered a valid benchmark assessment of mathematical skills of prospective tertiary students. The benchmark status of any assessment can only be measured against its intended goal. As has already been mentioned, MATs elect to focus on those aspects of the school curriculum that have a greater bearing on *performance in first-year mathematics courses*. MAT's suitability as a benchmark tool can therefore be established by how predictive it is of first-year mathematics performance. The reconfirmed discrepancy between MAT and first-year mathematics performance raised the question of suitability of MAT as a benchmark tool. Since the NSC examination results also do not show a dependable predictive value of first-year mathematics performance, there needs to be a revisit of the assessment of mathematical skills of prospective tertiary students. Alternatively, the reasons for non-predictive value of NSC examinations need to be explored and remedial steps taken. Identified shortcomings of MAT also need to be addressed. This could pave way for a common assessment of mathematical skills for the prospective tertiary students.

References

- ASSOCIATION OF AMERICAN COLLEGES AND UNIVERSITIES. 2011. *The LEAP vision for learning: Outcomes, practices, impact, and employers' view*. Washington, DC.
- ATTAL, Y., & BURSTEIN, J. 2006. Automated essay scoring with e-rater. *Journal of Technology, Learning and Assessment*, 4(3).
- BINKLEY, M., ERSTAD, O., HERMAN, J., RAIZEN, S., RIPLEY, M., & RUMBLE, M. 2012. Defining 21st century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills*. (pp 17 – 66). New York, NY: Springer Science and Business Media B.V.



- BUTLER, H.A. 2012. Halpern Critical Thinking Assessment predicts real-world outcomes of critical thinking. *Applied Cognitive Psychology*, 25(5): 721 – 729.
- CHALLIS, D. 2002. Integrating the conceptual and practice worlds: A case study from architecture. In A. Goody, J. Herrington & M. Northcote (Eds), *Quality conversations: Research and Development in Higher Education*, 25:106–113).
- COLIN, P., & WREN, J. 2005. Exploring Reliability in Academic Assessments. Retrieved from <http://www.uni.edu/chfasoa/reliabilityandvalidity.htm>
- EDUCATIONAL TESTING SERVICE. 2010. *ETS Proficiency Profile user's guide*. Princeton: NJ.
- HALPERN, D.F. 2003. *Thought and knowledge: An introduction to critical thinking*. Mahwah, NJ: Erlbaum.
- KLEIN, S., LIU, O.L., SCONING, J., BOLUS, R., BRIDGEMAN, B., KUGELMASS, H., NEMETH, A., ROBBINS, S., & STEEDLE, J. 2009. Test Validity Study (TVS) Report. (September 29, 2009).
- LEACOCK, C., & CHODOROW, M. 2003. C-rater: Scoring of short-answer questions. *Computers and the Humanities*, 37:389-406.
- LIU, O.L., FRANKEL, L., & ROOHR, K.C. 2014. Assessing Critical Thinking in Higher Education: Current State and Directions for Next-Generation Assessment. Research Report ETS RR-14-10.
- LIVINGSTONE, S.A. 2009. Constructed-Responses Test Questions: Why Use Them; How We Score Them. ETS R & D Connections – 09–11.
- MATZEN, R. N. Jr., & HOYT, J. E. (2004). Basic writing placement with holistically scored essays: Research evidence. *Journal of Developmental Education*, 28(1): 2-4,6,8,20,23,34.
- NBT. 2013. www.nbt.ac.za/content/preparing-your-learners-0.
- SWARTZ, S. M. 2006. Acceptance and Accuracy of Multiple Choice, Confidence-Level, and Essay Question Formats for Graduate Students. *Journal of Education for Business*, 81(4):215-220.