Per Linguam 2020 36(2):76-89
http://dx.doi.org/10.5785/36-2-885

Journal Home Page:
http://perlinguam.journals.ac.za

Per Linguam

A Journal for Language Learning
Tydskrif vir Taalaanleer

VALIDATING THE HIGHEST PERFORMANCE STANDARD OF A TEST OF ACADEMIC LITERACY AT A SOUTH AFRICAN UNIVERSITY

Kabelo Sebolai & Fiona Stanford
Language Centre, Stellenbosch University

ABSTRACT

Poor graduation rates are a serious concern worldwide. In South Africa, this concern has escalated in the post-apartheid era wherein a democratic constitution has widened access to higher education for school-leavers. The socioeconomic and school backgrounds of the majority of these learners still hamper their timely completion at university. In order to combat this, local universities have implemented some necessary interventions. Such interventions are geared towards dealing with the academic language needs of incoming students. For the last two decades or so, standardised tests of academic language ability, now commonly known as academic literacy, have been used to determine these needs. Given the expected impact of these interventions on student completion rates, the importance of the validity of these tests cannot be overemphasised. The aim of this article is to investigate the validity of the highest performance standard set for one of the tests currently used to assess levels of academic literacy. Using 14 610 scores obtained on that test by first-year students at a South African university, in tandem with their average scores on completion of their first year, sensitivity and specificity statistics were computed to realise this aim. The results revealed that the performance standard investigated was valid 61% of the time.

Keywords: National Benchmark Test in Academic Literacy, validity, performance level, sensitivity, specificity, classify

INTRODUCTION

Worldwide, the last 20 to 30 years have seen a significant increase in the number of students gaining admission to universities. In South Africa, this massification has been accelerated by the increased access of students from low socioeconomic backgrounds since the onset of the post-apartheid period. This is evident, for example, in a case study carried out on this situation at South African universities in 2001 and 2010, which reported a 'massive increase' in enrolments by African and coloured students (Wingate, 2015). Regardless of its expected long-term positive outcome, this development has also created a challenge for the higher education sector. The issue is that the retention and graduation rates of these students have been disappointingly low, with reports (e.g. Higher Education Funding Council for England, 2013) showing, for example, that of the student cohort that enrolled in 2004, only 38.3% of the African students and 42.1% of the coloured students had graduated by 2009, as opposed to 63.5% of the white students. This indicates that two of the student groups are likely experiencing the most difficulty overcoming the articulation gap, a problem experienced by most South African students upon admission to university. The articulation gap has been

described as the mismatch between the skills that students leave high school with as opposed to what they are expected to be able to accomplish when they reach university.

Among the identified sources of this articulation gap is a student's inability to cope with the language demands of academic education, a competence now commonly known as academic literacy. Currently, discipline-specific academic literacy is favoured over the generic approach which prevailed in the past. In accordance with this more focused perspective, Wingate (2015) gives a succinct definition of academic literacy as 'the ability to communicate competently in an academic discourse community'. Yeld and Cliff (2006: 19) provide a more detailed view of this competence in the following words:

> students' capacities to engage successfully with the demands of academic study in the medium of instruction of the particular study environment. In this sense, success is constituted of the interplay between the language (medium of instruction) and the academic demands (typical tasks required in higher education) placed upon students.

Similarly, other scholars, such as Boughey and McKenna (2016: 5), have defined academic literacy as 'the ability to read and write in socially legitimated ways in the academy'. In some quarters, this has been interpreted as the ability to 'integrat[e] content and language' (ICL) or as Jacobs (2013:135) puts it, 'providing access to knowledge through language'. Jacobs (2013: 132) further argues that if students are to be successful in their various fields of study, then 'what counts as knowledge in the discipline [should be made] explicit [to create] new knowledge'. After all, knowledge within a discipline is what is assessed in examinations throughout a student's years of study. Furthermore, if knowledge is central to academic literacy, then this could indicate a move away from focusing on teaching language *per se*, although language remains an important means of providing access to knowledge within a particular field (Van Rooy & Coetzee-Van Rooy, 2015). Finally, Freebody, Maton and Martin (2008: 196) add that students 'need to learn the reading, writing, talking and listening rules of the game for each subject area if they wish to succeed'. It is from this discipline-specific point of view that several scholars such as Gee (2012) and Boughey and McKenna (2016) have argued that what has until now been known as academic literacy be broadened and called 'literacies'. This is so as to do justice to the complexity of this competence by including aspects such as reasoning, critical thinking, deep-level text analysis, as well as academic acculturation, which are all important for academic success (Boughey & McKenna, 2016).

The different dimensions of the discipline-specific perspective mentioned in the previous paragraph all culminate in the assertion that language is an essential tool for successful academic performance. This means that students with linguistic or educational backgrounds that do not conform to the required literacy conventions of university study are often held back by this. This makes sense particularly when one considers Bourdieu and Passeron's (1990) view that academic language is no one's mother tongue, and that the articulation gap with regard to language is perhaps less about student diversity and more about the differences between and demands of specific disciplines as well as academic acculturation (Boughey & McKenna, 2015; Van Dyk, 2015).

The value attached to discipline-specific definitions of academic literacy notwithstanding, standardised discipline-specific tests of academic literacies are yet to be developed and used by South African universities for the placement of students within specific disciplinary programmes. Such tests are, firstly, crucial for determining the degree of academic

preparedness among incoming students and the degree to which a lack thereof will hinder their academic achievement. Secondly, they are important for ensuring that any support provided to empower students is sufficiently tailored to meet the demands and challenges they will encounter during their years at university (Cliff, 2015; Van Dyk, 2015). The non-existence of discipline-specific tests has meant, however, that South African universities continue to use the only two generic tests of academic literacy that are currently available to them. This makes it necessary for these tests to be continuously investigated for the utility of the information they provide about the extent of the articulation gap among the students that gain admission to academic education. This information is, in the case of these tests, determined and reported through performance levels or standards that are delineated by cut scores or benchmarks.

The aim of the present article is to investigate the extent of accuracy with which the highest benchmark or cut score for the test known as the National Benchmark Test in Academic Literacy (NBT AL) classifies test takers. In other words, the focus of the article is to investigate the validity of the highest benchmark that is set for this test. As will be explained later, this is the category where, according to the owners of the test, students whose scores fall within or above it will be unlikely to need academic support in order to successfully complete their studies. The reason for focusing on this particular performance level or standard is, as will be shown below, that the largest proportion of the participants in the study tended to score within this level of performance. Therefore, the size of the sample constitutes a strong basis for making sure that the results of this study are valid. In order to provide the correct context for this article, we focus on the meanings of the term validity and of benchmarks, cut scores or performance standards below. This is preceded by a brief review of the literature relevant to this article.

LITERATURE REVIEW

Although the NBT project has existed for longer than two decades to date, the literature focusing particularly on the relationship between its tests and academic performance is very limited. This is notwithstanding the fact that academic performance is supposedly the criterion informing the constructs of these tests. For the purpose of locating this article in this literature, three studies of immediate relevance to the article are worth brief exploration. The first was carried out by Van Rooy and Coetzee Van Rooy (2015) and focused on the relationship between several measures of language ability, which included the NBT AL, on the one hand, and first-, second- and third-year academic performance at the North West University on the other. Van Rooy and Coetzee Van Rooy (2015) used correlational analysis as the methodology for their study. One of their findings was that this test and other independent variables involved were not good predictors of academic performance. Van Rooy and Coetzee Van Rooy (2015: 42) summarise this finding as follows: 'One of the most important findings to consider is that these participants' academic success at university was not predicted very strongly by language-related measures such as achievement in the NBT, TALL/TAG or even matric language results.'

The second study was by Sebolai (2016), which also, among other independent variables, investigated the predictive validity of the NBT AL at the Central University of Technology. Sebolai (2016) used the regression methodology to investigate both the predictive and incremental validity of all the independent variables he investigated. Sebolai (2016) found that other predictor variables, which included Grade 12 English results and another test of academic literacy, known as the Test of Academic Literacy Levels (TALL), positively

predicted first-year academic performance in that context, while NBT AL did not. A limitation of Sebolai's (2016) study was that the size of his sample was not as large as it was in the case of the Van Rooy and Coetzee Van Rooy's (2015) study.

The third was a study by Sebolai (2019) which focused on the relationship between, among others, performance within the two performance standards set for the NBT AL, namely Intermediate and Proficient, on the one hand, and the end of first year academic performance on the other. The sample for this study was drawn from the first-year cohort of students in several faculties at Stellenbosch University. Sebolai (2019) used both the correlational and analysis of variance methodologies to determine if there was any positive relationship between the two independent variables and the dependent variables, and whether the independent variables, namely the Intermediate and Proficient standards of the test, did in fact classify students as purported by its owners. The focus of this study was, in other words, to validate the NBT AL at the level of performance standards. This makes Sebolai's (2019) study the most relevant to the present article. The study found that indeed, students who performed within the highest performance standard of the test tended to perform better in their programmes of study on average at the end of their first year when compared to those who performed within the lower standard of performance set for the test.

THE CONCEPT OF VALIDITY

To date, the term validity has been defined in two ways. The first definition is that it is a measure of whether a test is able to satisfy the purpose for which it is intended. The second is that validity is a function of how test scores are interpreted and used. The latter definition is problematic when one considers that it separates test scores from the very test that generates them. This in itself leaves the purported difference between test and score validity, vulnerable. The approach in this article is therefore that tests are considered valid but that the scores they generate are only valid as a function of the validity of such tests. From the point of view of this approach, validity has been classified into three types that can broadly be associated with the internal and external dimensions of a test. As its name suggests, internal validity is concerned with the construct or trait or ability that underpins a test and covers all aspects of its content. External validity, on the other hand, covers the predictive, concurrent and, occasionally, the consequential validity of a test. On the one hand, construct validity is a term used with reference to a test owner's ability to show that the ability or trait that a test is intended to measure is theoretically defensible while the content type relates to the relationship between test content and the performance the test taker is expected to demonstrate in real life. On the other hand, criterion-related validity is a function of how performance on a test can be shown to correlate with another measure or criterion, external to that test. This kind of validity can either be determined by administering the two measures at around the same time, or by using performance on one of the measures to predict performance on the other, at some time in the future. These are known as concurrent and predictive types of criterion-related validity, respectively. It is the latter kind of criterion-related validity that the present study sought to investigate with regard to a cut score, benchmark or performance standard set for the NBT AL. A brief description of all the cut scores set for this test is presented below.

BENCHMARKS OR CUT SCORES FOR THE NATIONAL BENCHMARK TEST IN ACADEMIC LITERACY

Cut scores or benchmarks, as they are called for the NBT AL, are levels of performance that are set for a test to classify test takers with regard to their levels of achievement on the criterion informing that test's construct. A cut score is, in other words, 'a reference point, usually numerical […] used to divide a set of data into two or more classifications' (Cohen & Swerdlik, 2010: 6). As indicated in some detail in the following section, the criterion informing the NBT AL is academic readiness with regard to language. This means that in the case of this test, cut scores or benchmarks are indications of whether the test taker is adequately equipped to bridge the articulation gap and the extent to which this is the case. For the purpose of indicating the different levels of test taker standing with regard to this gap, performance on the NBT AL is classified into Basic, Intermediate and Proficient performance standards. As the names given to these classifications suggest, the Basic level of performance means that a test taker has the greatest articulation gap and that long-term support would need to be provided to those who fall within this level if they are to succeed at university. In the words of the developers of this test, it is predicted that those whose scores fall within this band 'will not cope with degree level study without extensive and long term support, perhaps best provided through bridging programmes […] or FET provisions' (Cliff, 2015: 18). The second, the Intermediate level, means that test takers whose scores fall within this band will need less support in order to succeed at university. However, it is predicted that 'academic progress will be adversely affected. If admitted, students' educational needs should be met as deemed appropriate by the institution' (Cliff, 2015: 18). The last, the Proficient level, means that a test taker whose score places them within this level can gain straight admission to an academic programme. Moreover, the student is likely to succeed without the kind of support recommended for those in the other two levels. Performance at this level suggests, in other words, that 'future academic performance will not be adversely affected' and that 'if admitted, students may be placed into regular programmes of study' (Cliff, 2015: 18).

For the NBT AL, these benchmarks are a result of a standard-setting process that is carried out every three years. The current benchmarks for degree-level study for this test are 38% and 64%. The former separates the Basic and the Intermediate levels, while the latter separates the Intermediate and Proficient levels of performance. This is captured in Table 1 below.

Table 1: The performance standards/levels for the National Benchmark Test in Academic Literacy

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. Degree: AL [64%]; QL [70%] MAT [68%] Diploma/Certificate: AL [64%]; QL [63%] MAT [65%]

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). Degree: AL [38%]; QL [38%]; MAT [35%] Diploma/Certificate: AL [31%]; QL [34%] MAT [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.

(National Benchmark Tests Project, 2015)

Regardless of the method used for setting cut scores, it is crucial that the degree of validity of these scores be determined. This is particularly critical for a test like NBT AL, which is widely used by universities for making placement and access decisions. The consequences of these decisions seem far-reaching when one considers Bejar's (2008: 3) remarks that

> cut scores that do not represent intended policy or do not yield reliable classifications of students can have significant repercussions for students and their families; fallible student-level classifications can provide an inaccurate sense of an educational system's quality and the progress it is making towards educating its citizens.

This suggests that cut scores should be not only consistent with intended educational policy, but also psychometrically sound (Bejar 2008). As an outcome of the process that generates them, cut scores should, in other words, be subject to scrutiny to ensure that they satisfy all psychometric properties of measurement. This point is very well captured in the description of the standard-setting process by the National Council on Measurement in Education (2010: 15):

> Standard setting is more appropriately conceived of as a measurement process. The construct being measured is the panellists' representation of student performance that is at the threshold of an achievement level, e.g., barely proficient or barely advanced. The measurement of that construct results in cut points recommended by panellists. Because standard setting is a measurement process, standard setting results should be evaluated using the same expectations and theoretical frameworks used to evaluate other measurement processes in education such as student measurement.

What this means is that standard setting itself is a measurement procedure to which all design principles of measurement, such as reliability, validity and accountability, must apply. This is at the heart of the research question informing the study carried out in this article: Does the Proficient cut score or the highest standard set for the NBT AL classify test takers with a

reasonable degree of accuracy? In other words, to what extent is the Proficient cut score of the NBT AL valid in classifying test takers as those who are likely to succeed in their studies and those who are not? A brief description of this test is presented below.

THE NATIONAL BENCHMARK TEST OF ACADEMIC LITERACY

The NBT AL is a part of the battery of tests developed by the National Benchmark Tests Project, the brainchild of a union of vice chancellors of South African universities currently known as Universities South Africa. The decision to introduce this kind of testing was a result of widespread concerns about the evident failure on the part of the secondary schooling sector to prepare high school leavers adequately for higher education. The battery of tests includes two others, the Quantitative Literacy and Maths tests. Together with the NBT AL, these tests are intended to measure test takers' levels of readiness for academic education within the domains specified by the names of these tests (Griesel, 2006).

This means that, unlike the Quantitative Literacy and Maths tests, the NBT AL is intended to measure the levels of academic literacy among first-time entrants to higher education. The purpose of the test is, in other words, to 'assess the ability of first-year students to cope with the typical language-of-instruction, academic reading and reasoning demands they will face on entry to higher education' (Cliff, 2015: 4). Consequently, the construct that is measured by this test has been defined as a student's ability to do the following (Yeld & Cliff, 2006:20):

- negotiate meaning at word, sentence, paragraph and whole-text level;
- understand discourse and argument structure and the text 'signals' that underlie this structure;
- extrapolate and draw inferences beyond what has been stated in text;
- separate essential from non-essential and super-ordinate from sub-ordinate information;
- understand and interpret visually encoded information, such as graphs, diagrams and flow-charts;
- understand and manipulate numerical information;
- understand the importance and authority of own voice;
- understand and encode the metaphorical, non-literal and idiomatic bases of language; and
- negotiate and analyse text genre.

The NBT AL consists of 75 items in a multiple-choice format. All these items are developed by a team of experts from various fields in higher education and are reflective of the kind of tasks that students are likely to come across in their first year at university or college (Cliff, 2015).

SAMPLING AND DATA COLLECTION

The sample for this study comprised a total of 14 610 first-year students who were admitted to Stellenbosch University between 2013 and 2015. They were enrolled in the Faculties of Agriscience, Arts and Social Sciences, Economic and Management Sciences, Education, Engineering, Law, Medicine and Health Sciences, Science, and Theology. The Institutional Planning Department of the university provided all the data used for the study after institutional permission and ethical clearance had been obtained. These data consisted of the participants' scores on the NBT AL, the independent variable, and the average of their marks at the end of their first year, the dependent variable. Other variables, such as the participants'

demographics, were not considered as they were not of particular relevance to the present study.

METHODOLOGY

The methodology used for data analysis in this article involved computing two types of statistics known as sensitivity and specificity. The first refers to the degree of accuracy that a test possesses to predict accurately that test takers who perform well on that test will do the same on a future criterion, such as academic success, for example. In the words of Van der Walt and Steyn (2017: 109), 'sensitivity involves the number of true positives […], the proportion (or percentage) of students above a cut-off point who pass their first year'. The second statistic refers to the extent to which test performance can accurately predict that test takers who score low on the criterion informing the test will perform equally poorly in some future criterion that is related to the test. As Van der Walt and Steyn (2017:109) put it, 'specificity […] involves the number of true negatives […] the proportion (or percentage) of students below a cut-off point who fail their first year'.

In the case of this study, the focus is on the sensitivity and specificity of the highest performance standard set for the NBT AL, namely the Proficient band. In other words, the study set out to investigate how accurate the cut score or benchmark set for this standard was in classifying students who would obtain an average score of 50% for all their combined courses at the end of their first year of study and those who would not. Fifty percent was chosen as the reference point to determine this performance standard's degree of accuracy or validity because it is the minimum average score required to pass a course at most, if not all, South African universities. This means, by extension, that a student who enrols for five courses a semester and obtains 50% in all of them, for example, will pass and move on to the next level. Similarly, if this student obtains this score in all their courses throughout their studies, they will satisfy the minimum requirement for graduation. As was indicated earlier, the Proficient standard of NBT AL is set to classify those test takers who will obtain at least an average score of 50% in their programmes of study at the end of their first year.

RESULTS

Figure 1 below is a visual representation of how Stellenbosch University students typically perform on the NBT AL.

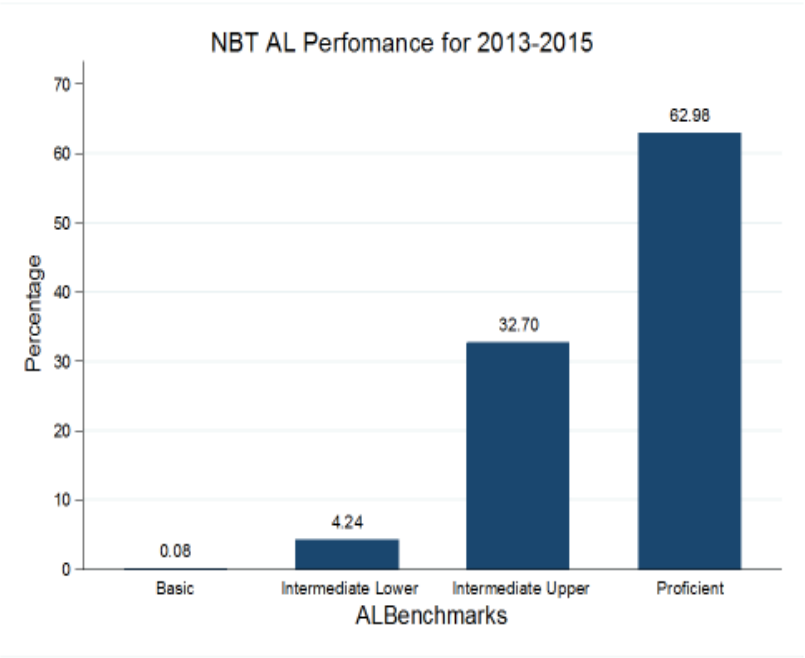

Figure 1: Typical performance by first year applicants to Stellenbosch University on the National Benchmark Test of Academic Literacy (n = 14 610)

As can be seen from this graph, the largest proportion of first-year applicants to Stellenbosch University tend to score within the Proficient band of this test rather than in its lower levels of performance.

Table 2, shown below, contains two sets of results that are the most relevant to the focus of the present article. The first is a result of cross tabulation of the scores in the two categories of performance chosen for this study. These are the average marks at the end of first year within the ranges 0-49 and 50-100, and those that fell within the three standards of performance set for the NBT AL, namely the Basic, Intermediate and Proficient bands. Secondly, the table depicts the results of a sensitivity and specificity analysis of performance within the three bands in relation to the end of first year average marks. Of interest to the current study in this set of results are the sensitivity and specificity values in the second and third columns for the Proficient standard as well as the 'correctly classified' value for the same standard.

Table 2: The results of a sensitivity and specificity analysis of the scores within the Proficient band

Year 1 Average	**Basic**	**Intermediate Lower**	**Intermediate Upper**	**Proficient**	**Total**
0-49%	5	279	1478	2023	3785
50-100%	6	340	3300	7179	10825
Total	11	619	4778	9202	14610

>= Cutpoint	**Sensitivity**	**Specificity**	**Correctly classified**	**LR+**	**LR-**
Basic	100.00%	0.00%	74.09%	1.0000	
Intermediate	99.94%	0.13%	74.09%	1.0008	0.4196

Upper					
Intermediate lower	96.80%	7.50%	73.67%	1.0466	0.4260
Proficient	**66.32%**	**46.55%**	**61.20%**	**1.2408**	**0.7235**
Greater than Proficient	0.00%	100.00%	25.91%		1.0000
ROC statistics					
Observations	**ROC area**	**Standard error**	**Asymptotic normal 95% confidence interval**		
14610	0.5696	0.0048	0.56022	0.57888	

As can be seen in Table 1 above, the sensitivity value for the Proficient standard equalled 66.32% while its specificity value was 46.55% if the average cut score for acceptable performance in the first year is 50%. Keeping the meaning of the concepts of sensitivity and specificity in mind, this means that the NBT AL Proficient band was able to classify the participants correctly as those that would score 50% and above on average at the end of their first year of study 66.32% of the time. It also means that this band was able to correctly classify those who would score below 50% on average 46.55% of the time. In other words, the first group included those whose score on the NBT AL was 64% (the current cut score for the Proficient standard) and above and who obtained 50% (the current cut score for a pass at Stellenbosch University) and above on average in their studies at the end of their first year. The second comprised those who scored below 65% in the test and scored below 50% on average in their studies at the end of their first year. Conversely, these results also mean that the Proficient band incorrectly classified 33.68% of the participants as those who would score 50% and above at the end of their first year of study and 53.45% as those who would score below 50% in their academic programmes at the end of the same year.

A combined statistic of the sensitivity and specificity of a measure is a final indication of its overall ability to classify test takers correctly. In the case of the Proficient standard of the NBT AL in this study, this amounted to 61.20%, as shown in Table 1 above. Put differently, the results of this study revealed that overall, the Proficient standard was able to classify the participants correctly as those that would score 50%, above 50% and below 50% on average at the end of their first year of study 61.20% of the time.

The graph shown in Figure 2 below is what is known as the receiver operating characteristic (ROC) curve that conventionally accompanies the results of a sensitivity and specificity analysis such as that captured in Table 1 above.

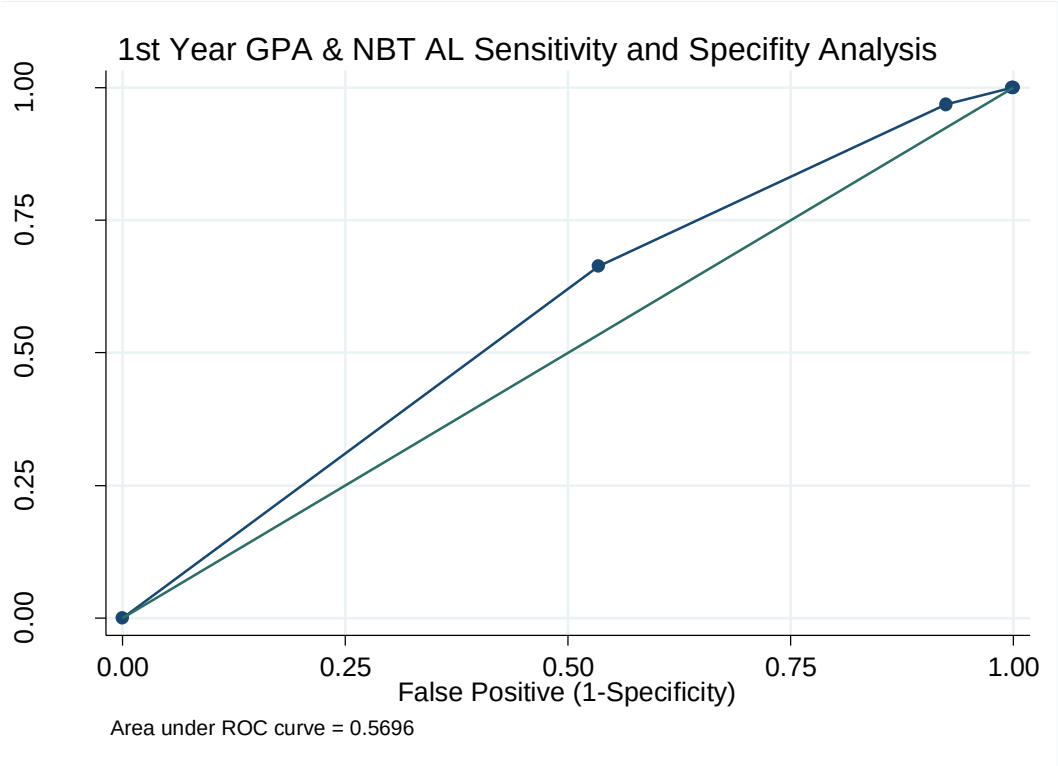

Figure 2: A visual plot of the ROC curve accompanying the results of the sensitivity and specificity analysis shown in Table 1.

In Figure 2 above, the point in the middle of the line above the one running diagonally across the graph and splitting it into two halves is the cut score set for the Proficient band. This is the ideal cut point for maximising true positives (sensitivity) and minimising false positives (1 – specificity). As can further be seen in the graph, the area between the diagonal line and the point at which the Proficient cut point is located equals 0.5696. If the size of this area, also known as the area under the curve (AUC), is greater than 0.05, it means that the NBT AL was, overall, a reasonably good predictor of the outcome variable, which was the end of first year average performance in the case of the present article (Van der Walt & Steyn, 2017). According to Van der Walt and Steyn (2007: 113), the 'Area Under the Curve is denoted by AUC and is used as *a measure of the ability to discriminate between the distributions* […] of the scores of the P and F populations. Larger values of AUC indicate a greater discrimination ability.' In the context of the study presented in this article, the P and F represent those who scored 50% and above and those who scored less than 50% at the end of their first year. In other words, the letters P and F denote Pass and Fail, respectively.

DISCUSSION

Firstly, the results of this study show that the majority of students admitted to Stellenbosch University possess high levels of academic literacy as measured by the NBT AL and that the minority tend to perform within the lower performance standards of this test. Not only is this evident in the graph showing this performance in Figure 1 above, but it is also evident in the results of a cross tabulation of the participants' end of first year average academic performance on the one hand and their performance within the three performance standards of the NBT AL on the other. From this cross tabulation, it is also clear that performance on

the test appears to relate positively to academic performance at the end of the first year. It is evident, for example, that participants who scored below 50% on average in their academic programmes were represented more significantly within the lower standards of NBT AL than those whose average scores were 50% and above. On this test in particular, this kind of performance is typical of students admitted to 'historically advantaged' South African universities such as Stellenbosch University and the University of Cape Town.

Secondly, the results of the sensitivity and specificity analysis carried out in this article mean that the Proficient standard of the NBT AL was stronger in indicating that the group of students involved were likely to pass their first year of academic study and weaker in identifying those who were unlikely to do so. While it is important that a test of this kind is extremely efficient in both these cases, it is even more crucial that its ability to separate the latter group out is the strongest. As a matter of logic, these are the students who stand to benefit from the academic literacy courses that most, if not all, South African universities have offered for more than two decades to date for the purpose of enhancing the kind of language ability that the NBT AL assesses and which has been proven to play a role in general academic performance and, ultimately, student throughput rates. What is undesired and what the analysis referred to above revealed is that the test favoured those students who performed proficiently in the test with regard to how it predicted their end of first year performance. It is important to emphasise, in the context of testing for academic preparedness, that a student's good performance in a test designed for this purpose should not mean that such a test is predictively biased in favour of that student and against their counterpart. To this end, test performance and test bias are two related but distinct aspects of educational measurement which should always be dealt with as such. Although carried out in a different context, Sebolai's (2018) study underlines the importance of this argument using data from the only other South African standardised test of academic literacy, which was referred to as TALL earlier in this article.

Finally, it is also worth mentioning that, as shown in Table 1, the data analysis for this study yielded statistical results for sensitivity, specificity and 'correctly classified' for the other performance standards set for the NBT AL. What is clear from this table, for example, is that the other standards scored higher in the 'correctly classified' column than their Proficient counterpart. This was to be expected because, as was revealed by the cross tabulation of all the performance categories dealt with in the study, the largest proportion of the participants fell within the Proficient standard while the smallest fell within the lower standards. The small sample size in these lower standards is likely to skew the results, and was likely the case in the present study. Consequently, a meaningful comparison of the sensitivity and specificity of these standards to those of the Proficient standard is rendered meaningless. The higher values scored in the Basic and Intermediate standards as compared to the Proficient standard on the ability to 'correctly classify' the participants overall are therefore more than likely a result of smaller participant numbers within those bands. This provides the reasoning for the focus of this article being solely on the degree of validity of the Proficient performance standard and not on the others.

CONCLUSION

The aim of this article was to investigate the validity of the cut score set for the Proficient standard of performance on the National Benchmark Test in Academic Literacy. This cut score has been interpreted by the owners of this test to mean that students whose scores on this test fall within this category are unlikely to struggle with the demands of higher

education, can be admitted straight into mainstream academic programmes and are likely to succeed without extra intervention. This investigation was necessary because, while the levels of performance set for this test are an outcome of a supposedly rigorous process of standard setting, no evidence of the validity of these standards has been published in the two decades since its introduction. This means that universities have been using the test for whatever purpose and adhering to the interpretation advanced for these standards without any knowledge of the degree of its validity for their contexts. This knowledge is very important for institutions making use of the test since it is often the basis for high-stakes decisions, and it is imperative that test score analysts are aware of the extent of error that is always present in measurement.

The results of the study presented in this article showed that the Proficient standard was reasonably valid in separating students who would obtain the required average score to pass at the end of their first year of study and those who would not. The results also showed that the test itself was a reasonably good predictor of academic success overall for the particular group of participants used in the study. The outcome variable used to determine this was the average academic performance of 50% at the end of the first year of study. This means that students could obtain this score as an average for all their courses, irrespective of whether they performed equally or unequally in those courses. It also means that such a student could perform extremely well in some modules and extremely poorly in others but still obtain an average score of 50% overall. This is one weakness of the study carried out in the present article worth acknowledging. Similar future research should focus only on students who passed all their individual courses and obtained an average mark of 50% or above at the end of their first academic year. This will contribute greatly towards knowledge about the validity of the NBT AL performance standard dealt with in this article. Finally, one should be careful not to generalise these results to other university contexts. This is especially true when one considers that, unlike the situation at several other universities in the country, students at Stellenbosch University tended to fall within the Proficient band over the three-year period covered in this article.

REFERENCE LIST

BEJAR, I. I. 2008. Standard setting: What is it? Why is it so important? *R & D Connections.* Number 7, October. Educational Language Testing (ETS), Princeton, USA.

BOUGHEY, C. & MCKENNA, S. 2016. Academic literacy and the decontextualized learner. *Critical Studies in Teaching and Learning*, 4(2):1-9.

BOURDIEU, P. & PASSERON, J.C. 1990. *Reproduction in education, society and culture*. Newbury Park: Sage.

CLIFF, A. 2015. The National Benchmark Test in Academic Literacy: How might it be used to support teaching in higher education? *Language Matters*, 46(1):3-21.

COHEN, R.J. & SWERDLIK, M.E. 2010. *Psychological testing and assessment: an introduction to tests and measurement.* New York: McGraw-Hill.

COUNCIL ON HIGHER EDUCATION (CHE). 2010. Higher Education Monitor: Access and throughput in South African Higher Education: Three case studies. *HE Monitor Number 9*. Pretoria: CHE.

FREEBODY, P., MATON, K. & MARTIN, J.R. 2008. Talk, text, and knowledge in cumulative, integrated learning: A response to 'intellectual challenge'. *Australian Journal of Language and Literacy*, 31(2):188–201.

GEE, J.P. 2012. *Social linguistics and literacies: Ideology in Discourses*. Fourth Ed. London: Routledge.

GRIESEL, H. 2006. The context of the National Benchmark Tests project, in H. Griesel (ed) *Access and entry-level benchmarks: the National Benchmark Tests Project*. Pretoria: Higher Education South Africa.1 – 6.

HEFCE (Higher Education Funding Council for England). 2013. *International research on the effectiveness of widening participation* [Online]. Available: https://www.hefce.ac.uk/pubs/rereports/year/2013/wpeffectiveness [2017, December 2].

JACOBS, C. 2013. Academic literacies and the question of knowledge. *Journal for Language Teaching*, 47(2): 127-139.

NATIONAL BENCHMARK TESTS PROJECT. 2015. Standard Setting Workshop. Unpublished Information Pack. Cape Town: Higher Education South Africa.

NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION (NCME). 2010. Standard-setting methods as measurement processes. *Educational Measurement: Issues and Practice*, 29(1): 14-24.

SEBOLAI, K. 2016. The incremental validity of three tests of academic literacy in the context of a South African university of technology. PhD thesis. Bloemfontein: University of the Free State. Available: http://hdl.handle.net/11660/5408.

SEBOLAI, K. 2018. The differential predictive validity of a test of academic literacy for students from different English language school backgrounds. Southern African Linguistics and Applied Language Studies. DOI:10.2989/16073614.2018.1480899.

SEBOLAI, K. 2019. Validating the performance standards set for language assessments of academic readiness: The case of Stellenbosch University. *Stellenbosch Papers in Linguistics Plus*, 56, 79-95.

STEYN, HS AND VAN DER WALT, J. L. 2017. Setting a cut-off score for a placement test at tertiary level. *Journal for Language Teaching*. 51(2): 105-118.

VAN DYK, T. 2015. Tried and tested: Academic literacy tests as predictors of academic success. *Amsterdam University Press*, 37(2):159-186.

VAN ROOY, B. & COETZEE-VAN ROOY, S. 2015. The language issue and academic performance at a South African University. *Southern African Linguistics and Applied Language Studies*, 33(1):31-46.

WINGATE, U. 2015. *Academic literacy and student diversity: The case for inclusive practice*. Bristol: Multilingual Matters.

YELD, N. & CLIFF, A. 2006. Test domains and constructs, in H. Griesel (ed) *Access and entry-level benchmarks: the National Benchmark Tests Project*. Pretoria: Higher Education South Africa.17 – 27.

BIOGRAPHICAL NOTES

Kabelo Sebolai is the deputy director in the Language and Communication Development section of the Language Centre at Stellenbosch University. His professional background is Teaching English to Speakers of Other Languages (TESOL). His research interest revolves around academic literacy teaching and assessment. Email: ksebolai@sun.ac.za

Fiona Stanford is a lecturer at Stellenbosch University's Language Centre where she designs, co-ordinates and presents courses in academic literacy at various faculties within the University. Her research foci is in the field of language assessment and testing and the subject of her Master's degree was research into academic listening, and how it relates to academic literacy as a whole, within a university context. Email: fcm@sun.ac.za