

## Validating the performance standards set for language assessments of academic readiness: The case of Stellenbosch University

---

Kabelo Sebolai 

Language Centre, Stellenbosch University, South Africa  
E-mail: ksebolai@sun.ac.za

### Abstract

Twenty-five years into the post-apartheid period, South African universities still struggle to produce the number of graduates required for the country's socio-economic development. The reason most often cited for this challenge is the mismatch that seems to exist between the knowledge that learners leave high school with, and the kind that academic education requires them to possess for success. This gap, also known as the "articulation gap", has been attributed to, amongst others, the levels of academic language ability among arriving students. The school-leaving English examination, and a pre-university test of academic literacy are the commonly used measures to determine these levels. The aim of this article is to investigate whether predetermined standards of performance on these assessments relate positively with academic performance. In order to determine this, Pearson Correlations and an Analysis of Variance (ANOVA) were carried out on the scores obtained for these assessments by a total of 836 first-year students enrolled at Stellenbosch University. The results show that the performance standards set for the standardised test of academic literacy associate positively with first-year academic performance, while the scores on the levels of performance set for the school-leaving English examination do not.

**Keywords:** English, home language, English FAL, performance levels, validating, academic performance

### 1. Introduction

Over the last two decades, concerns over low graduation rates have been prominent in the debates in South African higher education. This is understandable as not only do low graduation rates affect the educational success of a country, they are also a measure of a country's socio-economic progress and standard of living. At the centre of these debates have been questions around the causes of these low graduation rates, and how best this can be dealt with. In general, everybody agrees that there is a gap in what high school graduates leaving this sector of education are able to do, and what they are expected to be competent in in order for them to succeed at university. This gap has invariably been associated with the socio-economic, emotional, and academic backgrounds of these school-leavers, with those with a political history of disadvantage being the most adversely affected. Universities have responded to this

by providing financial, wellness, and academic support to help these students bridge this gap. Amongst others, the academic support referred to above has involved academic language interventions that have often been preceded and, in some cases, informed by performance on the school-leaving English examination as well as a standardised test of academic literacy, namely the “National Benchmark Test in Academic Literacy” (NBT AL). The English examination is part of the school-leaving assessment that has traditionally been used by universities to make admissions decisions, while the NBT AL is designed to provide additional information about the extent of the gap, also known as the “articulation gap”, referred to above. On both these assessments, performance is classified according to different levels which should logically be interpreted in terms of the probability that those school-leavers/students who perform at higher levels will have a greater chance of performing well in their first year of academic study than those who perform at lower levels. More specifically, school-leaving English is taught and assessed at the Home Language (HL) and First Additional Language (FAL) levels, while performance on the NBT AL is reported at the Basic, Intermediate, and Proficient levels. The aim of this article is to determine whether these performance levels are consistent with the corresponding levels of academic performance that they suggest. The research question for this article is, in other words, twofold: Do the students who take English HL at school perform better in their first year of academic study than those who take it at the FAL level? Do those who are classified as proficient by the NBT AL perform better in their studies than those whose performance falls within the Intermediate band? The aim of the study is, in the final analysis, to determine not necessarily the validity of overall performance on these assessments, but rather that of their standards/performance levels.

Answering these two questions is important for two reasons. The first is that test-performance levels or standards are an outcome of standard setting, a measurement process of which the results, as the National Council on Measurement in Education (NCME 2010: 15) rightly points out, “should be evaluated using the same expectations and theoretical frameworks used to evaluate other measurement processes in education such as student measurement”. This is particularly necessary when one considers that, like those taken based on high stakes test scores, the decisions based on performance standards have consequences for both individual students and educational systems (Bejar 2008). The second reason is that, in the South African higher education context at least, studies focusing on the relationship between performance on the assessments dealt with in this article and academic performance have so far mainly approached this relationship from the angle of overall student performance, and not necessarily from the angle of the performance standards set for those assessments. Details of these studies are briefly laid out in the literature review (section 3) below. Prior to this review of the literature, however, a brief engagement with the meaning of the term “validity” is worth pursuing first.

## **2. Test validity**

The concept of ‘validity’ is at the centre of quality assurance in both language testing and applied linguistics research. Its meaning has to date been the subject of debate for over a century. So much has been written about this debate (e.g., Messick 1989; Bachman and Palmer 1996; Kane 2006, 2009; Sireci 2009; Newton and Shaw 2014) that it is not necessary to dwell in detail on the literature focusing on these contestations in this article. It suffices to point out that the debate has resulted into two schools of thought, the first one of which – the “traditional view” – regards validity as a function of how well a test measures what it purports to measure. This view suggests that validity is a property of a test, while the second and opposite view

maintains that validity resides in how test scores are interpreted and used. Furthermore, the traditional view classifies validity into the construct, content, and criterion-related types. The first type, construct validity, refers to the test developer's ability to justify the existence of the ability measured by a test based on a theory of some sort. The second, content validity, relates to the extent to which a test's tasks are a mirror of what the test-taker will be expected to perform in a real-life situation – the “Target Language Use domain”, as Bachman and Palmer (1996) call it. The third type, criterion-related validity, refers to the extent to which a test's results resemble those of another test measuring the same or a related construct. This can be determined using two tests that are administered concurrently or at two different times with a reasonable time lapse in between. The former method of validation results in what is known as the “concurrent type” of criterion-related validity, while the outcome of the latter is the “predictive type” of the same category of validity.

This categorisation of validity is, however, inconsistent with the second view of validity referred to earlier wherein test-score interpretation and use are the essence of the meaning of this concept. The latter view unifies validity into construct validity, and regards the traditional content and criterion-related types as mere sources of evidence for the former. As will become clearer later in this article, the criterion-related type of validity or evidence for construct validity is what this article seeks to investigate.

Weideman (2019a) has attributed the continuing controversy over the meaning of the concept of ‘validity’ to a lack of conceptual clarity. He suggests that instead of lumping the latter principles all under “validity”, such principles should be conceptualised as disclosures of that concept. This, Weideman (2019a, 2019b) argues, would mean that what he calls the “retrociatory and anticipatory principles” of language test design are distinguished from each other. In the words of Weideman (2019a: 6), the former “are constitutive, founding (‘necessary’) concepts, and include conceptual primitives relating, for example, to the reliability, validity, differentiation, intuitive appeal (“face validity”) and theoretical defensibility or construct validity” of a test, while the latter are regulative ideas that include “the technical significance, efficiency or utility” of that test. Weideman’s (2019a, 2019b) question is, in other words, whether the current tendency to subsume all these concepts under “validity” is not to overstretch it. Eventually, Weideman (2019a) suggests that the concept of ‘test validation’ should be replaced by “responsible test design”, which would mean that what he calls the “retrociatory and anticipatory” concepts of test design are distinctly recognised and ultimately satisfied for the purpose of quality assurance in test development.

### **3. Literature review**

The first study of relevance to the background for this article was carried out by Van Rooy and Coetzee-Van Rooy (2015). This study focused on, amongst others, the predictive relationship between performance on the school-leaving English FAL examination and NBT AL as independent variables, and the end-of-first-year average academic performance as the dependent variable for students at the Vaal Triangle campus of North-West University. The methodology used for this study involved computing correlations of the scores between these variables as well as the effect size for those correlations. Van Rooy and Coetzee-Van Rooy’s (2015) study found that both the NBT AL and school-leaving English FAL results were weaker than the overall matric results in predicting the participants’ end-of-first-year academic performance. Van Rooy and Coetzee-Van Rooy (2015: 42) summarise the findings of this study as follows:

[...] academic success at university for these participants was not predicted very strongly by language related measures such as the NBT, [...] or even matric language results. The average or general matric achievement remained a stronger predictor of academic achievement at university than any of these measures.

Of further interest in the findings of this study is that although the overall school-leaving examination results showed evidence of being predictively stronger than the rest of the other independent variables investigated, this proved to be more the case for school-leavers with higher marks on this examination than it was for those whose marks were lower on the examination and who were therefore at risk. It is on the basis of this finding that Van Rooy and Coetzee-Van Rooy (2015) conclude that overall school-leaving results are not a useful measure for making student-admission decisions if used alone. What makes this study different from the one carried out in my article is that it did not necessarily seek to establish whether the standards of performance set for English HL and FAL, on the one hand, and the NBT AL Intermediate and Proficient bands, on the other, predicted average academic performance differently as purported by these levels.

The next study also worthy of mentioning as part of the background to my research for this article was by Fleisch, Schöer and Cliff (2015). Its main focus was to determine if students coming from English HL and FAL school backgrounds were similar or different with regard to the levels of academic literacy they bring to university as measured by the NBT AL. The sample for this study comprised 759 students entering the Bachelor of Education degree programme at the University of the Witwatersrand (Wits) in 2014. The research problem for this study was that when these students were admitted on the basis of the required level of overall performance in the school-leaving examination, no consideration of their school-leaving levels of English – i.e., whether HL or FAL – was made, nor was it considered whether this could impact these students' levels of academic literacy in English. The concern that Fleisch et al. (2015) raise is that students who took English at these respective levels could end up obtaining the same overall score in the whole school-leaving examination when they could, in actual fact, be at different levels of academic literacy as a function of their different levels of English assessment in their school-leaving examination. This would mean that those who come from a FAL school background are less able to cope with the language demands of higher education as measured by the NBT AL. The results of this study revealed that:

almost 25 per cent of the English FAL students performed in the Basic classification of the NBT AL, compared to less than 0.5 per cent of the English HL students. If one adds to the Lower Intermediate category, then close to 80 per cent of the English FAL students would be in need of extensive and long-term support, compared to roughly 24 per cent of the English HL students.

(Fleisch et al. 2015: 167)

Not only did Fleisch et al.'s (2015) study reveal that English FAL school-leavers perform poorly overall on the NBT AL when compared to their HL counterparts, it also showed that this too was the case at the level of the subdomains of academic literacy that constitute the construct of academic literacy as defined for this test. These subdomains include students' ability to recognise/understand/use cohesion, communicative function, essential versus non-essential information, grammar (syntax), inferencing, metaphorical expressions, discourse relations, text

genre, and vocabulary (see Cliff and Yeld 2006; Cliff 2014, 2015). With regard to differential performance between the two groups on the ability to work out the meaning of vocabulary in context, Fleisch et al. (2015: 170) found that the gap for the students whose marks in English HL and FAL were around 60% was less than .5 of the standard deviation, but that this gap widened as these marks increased. In other words, “for the cluster of questions that [required] students to possess a lexicon of vocabulary related to academic study or to be able to work out word meaning from context, there [was] consistently a substantial gap between students who wrote the HL examination and those who wrote the FAL examination” (Fleisch et al. 2015: 170). Fleisch et al.’s (2015: 170) study further revealed that this gap widened in the students’ performance on the items assessing knowledge of metaphorical language. This suggested that “inequality in achievement is exacerbated for higher-order academic literacy competencies, in this case ability to understand, or work out from context, which is the basis of analogous reasoning, understanding non-literal language, [and] understanding the sociolinguistic meanings of idiomatic language” (Fleisch et al. 2015: 170). In the final analysis, the finding of this study was that this gap was the widest in the two groups of participants’ performance on those subdomains of academic literacy that required the highest level of processing capacity. In the words of Fleisch et al. (2015: 172),

students who matriculated with English HL performed on average around half a standard deviation higher than students who matriculated with English FAL when it comes to understanding language cohesion, essential/non-essential, text genre and vocabulary. In competencies such as grammar (syntax), inferencing, metaphorical expressions, and relations discourse, the difference is almost one full standard deviation.

While these results show that the level of high-school English related with test-taker performance on the performance standards set for NBT AL, they do not automatically mean that this kind of relationship exists in the participants’ actual performance at university.

Another study involving performance by the two groups of participants used in this article was carried out by Sebolai (2016). Amongst others, this study focused on the ability of the school-leaving English examination and the NBT AL to predict first-year academic performance for students at a South African university of technology. The study found that the former was the best predictor of first-year academic performance, and that the latter was not. It also revealed that the NBT AL was not able to add to the predictive power of the school-leaving English examination. Sebolai (2016) cites several possible reasons for the NBT AL’s predictive shortcoming in his study, one of which was the possible discrepancy between the test’s level of difficulty and the ability levels of students enrolling at a university of technology on the basis of lower admission requirements when compared to those who gain admission to traditional academic universities where these requirements are higher. Like the other two studies presented above, Sebolai’s (2016) study focused on the overall ability of these assessments to predict first-year academic performance. It did not, in other words, separate the participants into those who performed within the different standards set for these assessments.

Sebolai (2018) went further by investigating whether another test of academic literacy – the “Test of Academic Literacy Levels” (TALL), commonly used alongside the NBT AL by South African universities – predicted first-year academic performance differentially for students from an English HL school background as opposed to those from an English FAL school

background at the same university of technology referred to in the previous paragraph. In essence, this study aimed to determine whether this test was biased in favour of the two groups of students in terms of its ability to predict their end-of-first-year academic performance in their various programmes of study. Sebolai (2018) asserts the value of the focus of his study by arguing that under-preparedness with regard to academic language is likely to be an issue among students from English FAL school backgrounds. Such students are, in Sebolai's (2018: 3) words, those "who are likely to need extra academic language support to give them a fair chance of success at university study. The tests employed to determine this should therefore not be less efficient in predicting the possibility that these students will need this support". The results of this study showed that there was no significant difference in how the TALL predicted academic performance for the two groups of students. However, as with those mentioned earlier, this study's focus was on the validity of this test at overall average score level, and not on such scores falling within a particular standard of performance set for the test.

The last study also worth a brief mention was by Myburgh (2015). It focused on the ability of two academic literacy tests that were specifically designed to assess the academic literacy levels among high-school learners and the mid-year English HL results to predict the end-of-year academic performance for Grade 10 learners at two schools in Bloemfontein. Myburgh's (2015) study arose from the fact that, while academic language ability was/is a purported component of the high school HL English curriculum of South Africa's basic education, questions had been raised about whether this was also the case with the assessment that accompanied this curriculum (see Du Plessis, Steyn and Weideman 2016). At the time of Myburgh's study, there was concern that the assessment following this curriculum did not show any evidence of focusing on academic language ability. Myburgh's (2015) study could therefore potentially help to pioneer the best way of assessing academic language ability at the Grade 10 level of education. As Myburgh (2015) puts it, if academic language ability was part of what the high school HL English curriculum aimed to develop, it was vital that the nature of the tests used to assess this ability was established. The results of Myburgh's (2015) study revealed that although the mid-year HL English examination was the best predictor of the outcome variable, one of the two tests of academic literacy was almost as effectively predictive as the June HL English marks. Like the other studies referred to above, Myburgh's (2015) focused on predictive validity as a function of overall average scores, and not necessarily on such scores as classified according to the performance standards set for the predictor variables involved.

#### **4. The school-leaving English examination**

The Curriculum and Assessment Policy Statement (CAPS) that governs the English HL and FAL examinations stipulates that the teaching which leads to this assessment should train learners to become role players in "society as citizens of a free country", have "access to higher education", and facilitate their transition from "educational institutions to the workplace" (Department of Basic Education 2011: 4). To this end, CAPS has specified the aims of HL and FAL learning as a learner's ability to do the following:

- Acquire the language skills required for academic learning across the curriculum;
- Listen, speak, read/view and write/present the language with confidence and enjoyment. These skills and attitudes form the basis for life-long learning;
- Use language appropriately, taking into account audience, purpose and context;

- Express and justify, orally and in writing, their own ideas, views and emotions confidently in order to become independent and analytic thinkers;
- Use language and their imagination to find out more about themselves and the world around them. This will enable them to express their experiences and findings about the world orally and in writing.
- Use language to access and manage information for learning across the curriculum and in a wide range of contexts. Information literacy is a vital skill in the ‘information age’ and forms the basis for life-long learning; and
- Use language as a means for critical and creative thinking; for expressing their opinions on ethical issues and values; for interacting critically with a wide range of texts; for challenging the perspectives, values and power relations embedded in texts; and for reading texts for various purposes, such as enjoyment, research and critique.

(Department of Basic Education 2011: 9)

In addition to this, CAPS aims to promote language-learning achievement at the social and academic levels of proficiency. With regard to the social area of proficiency, the focus is on “the mastery of basic interpersonal communication skills required in social situations”, whereas the focus on the academic level is the “cognitive academic skills essential for learning across the curriculum” and “literary, aesthetic and imaginative ability” (Department of Basic Education 2011: 9). It is in view of this that Du Plessis, Steyn and Weideman (2016: 7) observed that the construct of language ability that informs CAPS for languages “articulates the intention to develop learners’ differentiated language ability so that by the end of their school careers they have mastery of language(s) in a wide range of contexts and situations (educational and academic; aesthetic, political, economic; social and informational; ethical)”.

The school-leaving English assessment comprises an examination that is written at the end of the year as well as an oral assessment which is administered during the year. The written component of this examination takes the form of two papers that are assessed out of 250 marks in total. The first paper focuses on Language in Context, and consists of sections on Comprehension, Summary, and Language Structure and Conventions. The second paper focuses on literature wherein learners are assessed on their understanding of poetry, drama, and fiction. For this paper, the HL language students are required to answer questions on all three genres, while their FAL counterparts are required to engage with questions on any two of the three genres. The third paper focuses on the assessment of writing. For the HL candidates, the focus is on essay and transactional text writing, while their FAL counterparts are required to complete an essay, and longer and shorter transactional writing tasks. For both HL and FAL candidates, longer transactional texts comprise friendly/formal letters, formal and informal letters to the press, a curriculum vitae and covering letter, an obituary, an agenda and minutes of meeting, and a report. For the FAL group, shorter transactional text writing includes advertisements, diary entries, postcards, filling in forms, etc.

In the context of South African pre-university education, the distinction made between HL and FAL is unrelated to the languages that learners do or do not speak at home. Instead, it is a function of the level of difficulty or cognitive demand at which the teaching and assessment of a language is pitched (Department of Basic Education 2011). This means that the level or standards set for the English HL and FAL examinations dealt with in this article are different,

with the HL examination pitched higher than that of its FAL counterpart. Du Plessis et al. (2016: 5) explain it as follows:

[...] the CAPS document makes it clear that the distinction of Home Language applies to the *level* at which the language is offered rather than the language itself. The standard thus set for HL level is higher than that set for FAL level, although in practical terms the competency level of a learner may be the same for both levels.

It is on the basis of this distinction that the University of Pretoria considers the level at which incoming students passed school-leaving English useful for placing such students into academic literacy courses. For example, in the Faculty of Humanities at this university, students who scored at level 4 or lower on the school-leaving HL English examination are required to enrol for two academic literacy modules, whereas those whose English FAL scores are at level 5 or lower are also obliged to take those modules. Level 5 is a higher point of achievement than level 4. This is a clear indication of the perception in that faculty that HL students are at a higher level with regard to academic language readiness when compared with their FAL counterparts, and that the former should therefore perform better academically compared to the latter. As pointed out earlier, part of the aim of the study carried out in this article is to determine if this is indeed the case.

## 5. The National Benchmark Test in Academic Literacy

The NBT AL is part of a battery of tests that have, for the last two decades, existed under the auspices of the National Benchmark Tests Project (NBTP), the brainchild of Universities South Africa (USAF), a union of the vice-chancellors of South African universities. The project was mooted against the backdrop of low levels of graduation rates at the country's universities, and the resultant need to assess entry levels of academic readiness for higher education in the country. As its name implies, the NBT AL was designed specifically to assess levels of student readiness to engage successfully with higher education studies in the languages of teaching and learning. In the words of Cliff and Yeld (2006: 19), this test aimed to assess

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.

The construct underpinning the NBT AL is largely informed by the view of language ability espoused by Bachman and Palmer (1996) wherein language ability is understood to comprise what they call "language knowledge" and "strategic competence". As Bachman and Palmer (1996: 70) put it, language knowledge is a combination of organisational and pragmatic knowledge while strategic competence is a function of "a set of metacognitive components, or strategies, which can be thought of as higher order executive processes that provide a cognitive management function of language use, as well as other cognitive activities". To this effect, the NBT AL was designed to assess a student's ability to do the following:

- 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.

(Cliff and Yeld 2006: 20)

The NBT AL consists of a total of 75 multiple-choice items for which the test-taker is required to read generic texts that are typical of those that they will encounter and need to process at university. For each of these items, the students are required to “choose the most inclusive or plausible or reasonable answer from four options, where distractors have been specifically designed to be indicative of reading and reasoning misconception” (Cliff 2015: 11). To the extent determined by the developers of this test, these items are designed to assess levels of ability on the subdomains of academic literacy that have been specified from its construct and that were briefly referred to earlier in this article. The space allocation for each of these subdomains in the test is as follows: Essential/non-essential information (20%), Metaphor (15%), Inferencing (15%), Communicative Function (15%), Vocabulary (10%), Discourse Relations (10%), Sentence-Level Cohesion (5%), Grammar/Syntax (5%), and Text Genre (5%) (NBTP 2015).

Performance on all the tests developed by the NBTP comprises three levels, namely Basic, Intermediate, and Proficient levels. The meaning of each of these standards is captured in the table below (NBTP 2015: 17–18).

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.
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).
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.

As can be seen in this table, the descriptors for these bands suggest that those students whose scores are in the Proficient band stand a chance of performing better in their studies than those whose performance falls within the Intermediate and Basic bands. Part of the aim of this article, as pointed out in the introduction, is to determine if this is indeed the case with the Proficient and Intermediate bands specifically.

## 6. Description of the sample

The sample for the study reported in this article was conveniently selected among the groups of first-year students admitted to different programmes offered within the nine faculties of Stellenbosch University over a period of three years (2015–2017). The aim was to ensure that students from all of these faculties were reasonably represented in the total sample ultimately selected. These faculties and the sample sizes of the participants from each were as follows: Agrisciences (105), Arts and Social Sciences (105), Economic and Management Sciences (105), Education (105), Engineering (105), Law (83), Medicine and Health Sciences (100), Science (105), and Theology (23). Other demographics of the participants were not taken into account as they were not of particular interest to the study.

## 7. Methodology

The methodology used in this article comprised three steps. The first step involved computing descriptive statistics for the scores obtained within the two levels of performance set for school-leaving English, namely HL and FAL, on the one hand, and two of the three levels set for the NBT AL, namely the Intermediate and Proficient levels, on the other. These assessments were the two independent variables in the study. Alongside this, the descriptive statistics were also computed for the end-of-first-year average performance, the dependent or outcome variable in the study. The next step focused on computing a correlation of the scores obtained within the levels of performance of the two independent variables with those of the outcome variable, the end-of-first-year average performance. The last step was an Analysis of Variance (ANOVA) involving the performance-level scores on the two assessments, i.e., the school-leaving English examination and the NBT AL, in relation to the end-of-first-year average performance. The Statistical Package for Social Sciences (SPSS) was used for all data analyses.

## 8. Results

Presented in Table 1 below are the descriptive statistics of the participants' performance on the English HL examination in relation to their average performance at the end of their first year of study.

**Table 1:** Descriptive statistics for English HL and first-year-average scores (n=534)

	Mean	Std. Deviation	N
English HL	73.6161	7.20939	534
First-Year Average	58.5408	7.45272	534

As can be seen from the data in this table, the mean performance of the two variables differed markedly while the standard deviations of the scores obtained on them were negligibly different.

In Table 2 below, the results are captured of a correlational analysis of the participants' scores on the English HL examination and their end-of-first-year average scores.

**Table 2:** Correlation for English HL and first-year-average scores (n=534)

		English HL	First-Year Average
English HL	Pearson Correlation	1	.262**
	Sig. (2-tailed)		.000
	N	534	534
First-Year Average	Pearson Correlation	.262**	1
	Sig. (2-tailed)	.000	
	N	534	534

\*\* Correlation is significant at the 0.01 level (2-tailed).

As can be seen in this table, the correlation coefficient for the two assessments amounted to .262, while the *p* value equalled .000.

The descriptive statistics for the participants' performance on the English FAL examination in relation to their end-of-first-year average performance are captured in Table 3 below.

**Table 3:** Descriptive statistics for English FAL and first-year-average scores (n=302)

	Mean	Std. Deviation	N
English FAL	72.6026	8.40147	302
First-Year Average	57.8616	7.31302	302

From this table, it is evident that the mean performance by the participants on the two variables was as markedly different as it was for their English HL counterparts, and that the difference in the standard deviations for their scores on the same variables was also not large.

Table 4 below presents the results of a correlational analysis of the participants' performance on the English FAL examination and their end-of-first-year average scores.

**Table 4:** Correlation for English FAL and first-year-average scores (n=302)

		English FAL	First-Year Average
English FAL	Pearson Correlation	1	.359**
	Sig. (2-tailed)		.000
	N	302	302
First-Year Average	Pearson Correlation	.359**	1
	Sig. (2-tailed)	.000	
	N	302	302

\*\* Correlation is significant at the 0.01 level (2-tailed).

As can be seen in this table, the correlation coefficient for the two variables amounted to .359 with a  $p$  value of .000.

In Table 5 below, the results are captured of an ANOVA for the English HL and FAL participants' scores on their English examination in relation to their end-of-first-year average performance.

**Table 5:** The results of the ANOVA of the scores on the English HL and FAL examination in relation to the end-of-first-year average (n=835)

First-Year Average

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	88.998	1	88.998	1.624	.203
Within Groups	45701.984	834	54.799		
Total	45790.983	835			

This table shows that the  $F$  statistics yielded by the ANOVA amounted to 1.624 with its  $p$  value being .203.

Table 6 below presents the descriptive statistics for the participants' scores within the Intermediate band of the NBT AL in relation to their end-of-first-year average performance.

**Table 6:** Descriptive statistics for the NBT AL Intermediate and end-of-first-year average scores (n=363)

	Mean	Std. Deviation	N
NBT Intermediate	55.8457	6.17904	363
First-Year Average	56.9689	6.88224	363

As can be seen from this table, the difference was small between the mean scores obtained by the participants within the Intermediate performance level of the NBT AL and their standard deviations.

Table 7 below presents the results of a correlational analysis of the participants' Intermediate scores and their end-of-first-year average performance.

**Table 7:** Correlation between NBT AL Intermediate and end-of-first-year average scores (n=363)

		NBT Intermediate	First-Year Average
NBT Intermediate	Pearson Correlation	1	.111*
	Sig. (2-tailed)		.034
	N	363	363
First-Year Average	Pearson Correlation	.111*	1
	Sig. (2-tailed)	.034	
	N	363	363

\* Correlation is significant at the 0.05 level (2-tailed).

As is clear in the table, the correlation coefficient for the Intermediate NBT AL scores and the end-of-first-year average performance amounted to .111 with a *p* value of .034.

In Table 8 below, the descriptive statistics are presented of the participants' scores within the Proficient band of NBT AL and their end-of-first-year average performance.

**Table 8:** Descriptive statistics for the NBT AL Proficient and end-of-first-year average scores (n=473)

	Mean	Std. Deviation	N
NBT Proficient	72.4841	5.47222	473
First-Year Average	59.3135	7.63518	473

From Table 8 above, it is evident that the mean for the scores within the Proficient band was higher than that for the year-end average scores. It can also be seen from this table that the standard deviation for the end-of-year average was noticeably higher than it was for the NBT AL scores in the Proficient band.

**Table 9:** Correlation results for the NBT AL Proficient and end-of-first-year average scores (n=473)

		NBT Proficient	First-Year Average
NBT Proficient	Pearson Correlation	1	.160**
	Sig. (2-tailed)		.000
	N	473	473
First-Year Average	Pearson Correlation	.160**	1
	Sig. (2-tailed)	.000	
	N	473	473

\*\* Correlation is significant at the 0.01 level (2-tailed).

Table 9 above shows that the correlation coefficient for the scores in the Proficient band of the NBT AL, and the end-of-year average performance equalled .160 with a *p* value of .000.

Finally, the results in Table 10 are of an ANOVA for the scores obtained by the participants within the Intermediate and Proficient bands of the NBT AL with regard to the end-of-first-year average performance.

**Table 10:** The results of an ANOVA for the NBT AL Proficient and end-of-first-year average scores (n=835)

First-Year Average

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1129.071	1	1129.071	21.084	.000
Within Groups	44661.912	834	53.551		
Total	45790.983	835			

As shown in this table, the *F* statistic yielded by the ANOVA amounted to 21.084 with a *p* value of .000.

## 9. Discussion

The results of this study show that there is a positive association between performance within both the English HL and FAL examinations, on the one hand, and the end-of-first-year average performance, on the other. This is represented by the positive correlation coefficients between the two independent variables and the outcome variable. These coefficients are accompanied by  $p$  values that are lower than .05, the maximum value set for statistical significance in applied linguistics and second language research (see Mackey and Gass 2005, Dornyei 2007). This means that this relationship is probably real, and that it did not happen by chance. Interestingly, the correlation coefficient for English FAL and the end-of-first-year average was visibly higher than it was for its HL counterpart for the group of participants used in this study. Given the advantaged English language background that the HL group supposedly brings to university, one would expect that HL English would score the highest in terms of correlation with academic performance when compared to its FAL counterpart. However, as revealed by the results of this study, this was not the case. Equally interesting was the preliminary revelation by the descriptive statistical data analysis carried out in this study that the mean scores for both the English HL and FAL were marginally different, and that they were not in harmony with the mean scores for the end-of-year average performance. What was of ultimate interest to this study, however, were the results of the ANOVA carried out on these two assessments in relation to the end-of-year average performance. As shown in Table 5, the  $p$  value for the  $F$  statistic yielded by this analysis was higher ( $p = .203$ ) than the one set for statistical significance for this kind of study. This means that the difference in academic performance at the end of the first year of study between the HL and FAL groups was not significant in the statistical sense. This means, furthermore, that while the owner of these examinations places them at higher and lower levels respectively, in terms of cognitive demand, the difference is insignificant regarding how students from the two school language backgrounds perform at the end of their first year, at Stellenbosch University at least. This is, at the very basic level, initially signalled by the marginal difference in the standard deviations of the scores on these two levels of English assessment. Standard deviation is a measure of the average distance between the scores and the mean. In the context of pre-assessment for higher education, a standard deviation is an indicator of the extent to which a measure is able to spread test-taker scores out to facilitate decisions about their placement at different entry levels to academic study.

Additionally, the results of this study show that there was a positive relationship between performance within both the Intermediate and Proficient bands of the NBT AL, and the end-of-first-year average performance. Performance within the Proficient band related more strongly with the end-of-year average scores than the scores within the Intermediate band did with these average scores. Both these correlations yielded a  $p$  value of .000, meaning that they were both statistically significant. Furthermore, the results of an ANOVA on the scores within the two performance levels yielded an  $F$  statistic of which the accompanying  $p$  value was also statically significant. This means that, for the participants used in this study, performance within the Intermediate and Proficient bands related positively with how differently they performed academically at the end of their first year. In other words, students whose scores on the NBT AL fell within the Intermediate band performed less well than those whose scores were within the Proficient band. This shows that while the English HL and FAL levels of examination were unable to separate these participants into those who would probably do well and those who would not do well at university education, the NBT AL Intermediate and Proficient performance levels were indeed able to do so.

To this end, the NBT AL is, in the case of the participants used in this study, indeed designed to provide information about student readiness which the school-leaving English results could not. This information is particularly valuable for taking decisions about whether students should first be placed in extended-degree programmes or be admitted straight into mainstream academic programmes. These results are important in that while the NBTP founding document professes that the project aims to provide additional information to that generated by school-leaving results about student readiness for academic education, evidence for this has barely been forthcoming from the project to the users of these tests. Universities have tended to use the tests on the basis of merely what the owners of this project say about the utility of these tests. It should be noted, however, that the results of this study are neither necessarily generalisable to all first-year students across all years at Stellenbosch University, nor other contexts beyond that of this university. It is a well-known fact that student demographics in South Africa differ across the higher education sector every year, and that a sample of participants in different years and contexts may thus yield different results. I say this with full awareness that the NBTP may be using sample-independent models for developing their tests, and that this allows for test equivalence across different samples and contexts. However, this will need to be borne out by the results of further empirical and longitudinal investigation.

## **10. Conclusion**

The aim of this article was to investigate the validity of the levels of performance set for the school-leaving English language examination and a test of academic literacy used by South African universities for making student-admission and -placement decisions. More specifically, the article focused on determining whether the English HL and FAL school-leaving levels of examination, and the Intermediate and Proficient bands of the NBT AL related positively with academic performance. It also aimed to determine if these levels related to different degrees with the end-of-first-year academic performance as suggested by how performance within these levels is supposed to be interpreted. Both correlations and ANOVA were computed to determine these results. The results of the study revealed that, contrary to how their levels of performance are or should be interpreted, performance on the English FAL examination related more strongly with first-year average academic performance than its English HL counterpart did. In other words, while both these levels of school-leaving English assessments showed evidence of a predictive relationship with academic performance at first-year level, English FAL had the strongest relationship. These results further showed that the difference of this strength was not statistically significant, meaning that the two standards of this examination did not efficiently separate the two groups of students into those who would perform well or less well as a function of having taken English at the HL versus FAL level for their school-leaving certificate. Furthermore, the results showed that although there was also a positive relationship between performances within the two standards set for the NBT AL, this correlation was lower than it was for both HL and FAL English. In the case of the NBT AL, the Proficient band – the highest level of performance set for this test – possessed the strongest relationship than the Intermediate band, its lower-level counterpart.

Finally, the results showed that the extent to which these bands distinguished the students as those who would perform well as opposed to those who would do so less well was statistically significant. This means that, for Stellenbosch University students at least, the levels of performance set for this test showed evidence of being valid for how they are supposed to be

interpreted; those who were classified as Intermediate by the test performed less well academically when compared to those who were classified as Proficient by the same test.

## References

Bachman, L.F. and A.S. Palmer. 1996. *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.

Bejar, I.I. 2008. Standard setting: What is it? Why is it important? *R&D Connections* 7: 1–6. Available online: [https://www.ets.org/Media/Research/pdf/RD\\_Connections7.pdf](https://www.ets.org/Media/Research/pdf/RD_Connections7.pdf) (Accessed 2 July 2019).

Cliff, A.F. 2014. Entry-level students' reading abilities and what these abilities might mean for academic readiness. *Language Matters* 45(3): 313–324. <https://doi.org/10.1080/10228195.2014.958519>

Cliff, A.F. 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. <https://doi.org/10.1080/10228195.2015.1027505h>

Cliff, A.F. and N. Yeld. 2006. Test domains and constructs: Academic literacy. In H. Griesel (ed.) *Access and entry level benchmarks: The national benchmark tests project*. Pretoria: Higher Education South Africa. pp. 19–27.

Department of Basic Education. 2011. *Curriculum and assessment policy statement: Grades 10–12 English Home Language*. Pretoria: Department of Basic Education.

Dörnyei, Z. 2007. *Research methods in applied linguistics*. Oxford: Oxford University Press.

Du Plessis, C., S. Steyn and A. Weideman. 2016. The assessment of home languages in the South African National Senior Certificate examinations: Ensuring fairness and increased credibility. *LitNet Akademies* 13(1): 425–443.

Fleisch B., V. Schöer and A. Cliff. 2015. When signals are lost in aggregation: A comparison of language marks and competencies of first year university students. *South African Journal of Higher Education* 29(5): 156–178.

Kane, M.T. 2006. Content-related validity evidence in test development. In S.M. Downing and T.M. Haladyna (eds.) *Handbook of test development*. New York: Routledge. pp. 131–153. <https://doi.org/10.4324/9780203874776.ch7>

Kane, M.T. 2009. Validating the interpretations and uses of test scores. In R.W. Lissitz (ed.) *The concept of validity: Revisions, new directions, and applications*. Charlotte, NC: Information Age Publishing, Inc. pp. 39–64.

Mackey, A. and S.M. Gass. 2005. *Second language research: Methodology and design*. New York: Routledge.



Messick, S. 1989. Validity. In R.L. Linn (ed.) *Educational measurement*. Third edition. New York: American Council of Education/Collier Macmillan. pp. 13–103.

Myburgh, J. 2015. The Assessment of Academic Literacy at Pre-University Level: A Comparison of the Utility of Academic Literacy Tests and Grade 10 Home Language Results. Unpublished MA thesis, University of the Free State.

National Benchmark Tests Project. 2015. NBTP National Report: 2015 Intake Cycle – CETAP Report Number 1/2015. Unpublished report. 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.

Newton, P.E. and S.D. Shaw. 2014. *Validity in educational and psychological assessment*. Los Angeles: Sage.

Sebolai, K. 2016. The Incremental Validity of Three Tests of Academic Literacy in the Context of a South African University of Technology. Unpublished PhD dissertation, University of the Free State.

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* 36(2): 105–116. <https://doi.org/10.2989/16073614.2018.1480899>

Sireci, S.G. 2009. Packing and unpacking sources of validity evidence: History repeats itself again. In R.W. Lissitz (ed.) *The concept of validity: Revisions, new directions, and applications*. Charlotte, NC: Information Age Publishing Inc. pp. 19–37. <https://doi.org/10.1007/s11336-010-9180-6>

Van Rooy, B. and S. Coetzee-van Rooy. 2015. The language issue and academic performance at a South African university. *Southern African Linguistics and Applied Language Studies* 33(1): 31–46. <https://doi.org/10.2989/16073614.2015.1012691>

Weideman, A. 2019a. Degrees of adequacy: The disclosure of levels of validity in language assessment. *KOERS: Bulletin for Christian Scholarship* 84(1): 1–15. <https://doi.org/10.19108/koers.84.1.2451>

Weideman, A. 2019b. Validation and the further disclosures of language test design. *KOERS: Bulletin for Christian Scholarship* 84(1): 1–10. <https://doi.org/10.19108/koers.84.1.2452>