This is an accepted manuscript of an article published by the South African Association for Language Teaching in the *Journal for Language Teaching*, Volume 51, 2017, Issue 1, p. 271-295, available online at https://journals.co.za/content/journal/langt/browse

The refinement and uses of a test of academic literacy for Grade 10 students

Jo-Mari Myburgh-Smit
Research fellow, Unit for Language Facilitation and Empowerment, University of the Free State
Albert Weideman
Senior research fellow, Office of the dean, Humanities, University of the Free State

Abstract

The massification of higher education has led to a substantial increase in enrolments since 1993, and an astonishing 300% rise in first degree completion among black students. Yet questions remain about the level and adequacy of students' preparation at school for such study. Drop-out rates of learners remain unacceptably high both at school and university level. Language ability is often identified as being one of several hurdles that prevent success, especially in higher education. At school there is an apparent misalignment between the aims of the current Curriculum and Assessment Policy Statement (CAPS), and the subsequent instruction and assessment of students. CAPS requires that students should be prepared to handle academic discourse, yet no clear outline of what academic discourse entails is given. Consequently, many higher education institutions across the country require of students to write additional pre-admission or post-entry tests of language ability. In some cases the National Benchmark Test (NBT) is used to grant or deny access, or in others for placement of at-risk students on language development interventions, usually defined as "academic literacy" courses. The clear expectation is that these tests will have some measure of predictive value, or at least be useful as regards minimising risk of failure. Ideally, it would then be advantageous if students who need to improve their academic literacy levels could be identified at an earlier stage than university entry, whilst they are still in school. To monitor and gauge the value of language assessments and courses, however, one would first need appropriate, adequate and defensible assessment instruments. This paper discusses the need for and the refinement of an academic literacy test for Grade 10 students as a first step towards measuring and then developing the required level of academic literacy before entry into higher education.

Keywords: academic literacy; language development; higher education; access; language ability; language testing; language assessment

1. The national pre-tertiary and higher education context

This paper explores the relation among three different, yet connected, sets of considerations on which there is not yet sufficient consensus. The first set is made up of the impediments that stand in the way of fulfilling the expectations that tertiary institutions have of language instruction and development at school. The second set has to do with the effects on student preparedness of the massification of higher education over the last two decades, that often yields a diagnosis of language ability being the critical feature of such preparedness. Once that diagnosis, whether flawed or correct, has taken root in policy and administration at universities, the third and further questions are: (a) what kind of language intervention would be appropriate and effective to relieve the pressure on first-time students; and (b) how such underprepared students should be selected for placement on the planned interventions.

Taking the first set of considerations, we may observe that the current Curriculum and Assessment Policy Statement (CAPS) for Home Language contains the prescriptions to be used by South African teachers to guide their lesson planning, the execution of their language instruction, and their subsequent measurement of the language ability of their students. The measurement is conventionally administered in the form of post-instruction assessments, most notably in the nationally administered Grade 12 exit examinations. The curriculum remains the centrepiece in all of this, so its pronouncements and requirements are crucial. CAPS prescribes that students must be able to function competently within the following material lingual spheres (Weideman 2009:39) or discourse types (Department of Basic Education 2011):

- ✓ social (including inter-personal communication and the handling of information)
- ✓ economic/professional (including the world of work and commerce)
- ✓ academic (including academic and scientific language and advanced language ability for educational purposes)
- ✓ aesthetic (including language associated with the appreciation of literature and art)
- ✓ ethical (including an appreciation of the values embedded in language use) and
- ✓ political (including the critical discernment of power relations in discourse)

CAPS moreover not only refers to academic discourse as an essential kind of discourse for the high-level ability that the language curriculum sets as the general goal of language development, but notes that it is imperative for students to master academic discourse in order to be able to gain access to "further or Higher Education or the world of work" (Department of Basic Education 2011:9). CAPS appears to equate academic discourse with a "high standard of language" (Department of Basic Education 2011:9), which is the closest definition given of it in this policy document.

The vagueness of the definition makes its interpretation problematic: though they are certainly related, it is debatable whether a "high standard of language" and academic discourse are summarily interchangeable terms. Moreover, without a clear definition of academic discourse, teachers and students might remain at a loss about what it entails and how it can be assessed. That in turn questions the validity of the results obtained from the assessments of language ability that must follow the instruction based on this curriculum, since no clear construct has been articulated (Patterson & Weideman 2013:109). There is thus a need to explore in much greater detail a definition of academic discourse as a specific material lingual sphere (Weideman 2009:39).

Why attention to academic literacy levels is already important at school level is dependent not only on the curriculum requirements referred to above, but is also significant in view of historical developments in higher education, the second set of considerations referred to at the beginning. In its 25 April 2014 edition, *Rapport* reported that the number of black students who completed their tertiary education had increased by 300% since 1991 (Jeffery 2014). More recent statistics show that in the 20 years since 1996, higher education attendance per 100000 of the population grew by close to 445,5% or at a rate of 22,3% annually. Attendance increases have been driven by increases in population, and by increases in enrolment rates for the African population groups (Statistics South Africa 2017: 9). This shift from a type of elite education system to an education system which supports larger numbers of students was both foreseen and welcomed by the National Commission on Higher Education (NCHE) in 2001 (Department of Basic Education 2001; cf. too Department of Basic Education 2005).

Whilst in essence this is a good thing which many see as contributing towards "enhanced skills development for students, improved job and career opportunities, improvements in society, the economy and communities, and a commitment to realising the principles of life-long learning" (Cliff, Yeld & Hanslo 2003:1), it also brings with it its own challenges; we know, for example, that to be able to perform successfully at university, a student needs to be able to handle the kind of language used there: academic discourse. In a number of studies undertaken since the mid-1990s, it has become clear, however, that the ability of new entrants in Higher Education to handle academic discourse may not be at an adequate level (Van Rensburg & Weideman 2002:152). So we first need to ask whether the school curriculum places enough emphasis on the importance of teaching academic discourse in order to prepare learners for the demands of Higher Education, and second whether academic discourse is subsequently being assessed in a valid and responsible way. This is necessary, third, since students who come to university underprepared as to the language demands they will face there, need to be

responsibly, effectively and fairly identified in order to place them on the appropriate language development path, usually an academic literacy intervention.

This paper is therefore a contribution to the ongoing debate, also in the pages of this journal, that is enriching our understanding of how best to deal with the three sets of issues referred to above. It will examine how, given our current understanding, assessments of language ability should be employed, and what the limitations of their use are, before setting out the development and administration of a test of academic literacy at senior secondary school level, and its potential utility in contributing to both a much needed awareness of and a potential early solution to some of these concerns.

2. What language assessments and interventions can potentially predict, and be useful for

We take as the starting point of this discussion the argument of Van Rooy and Coetzee-Van Rooy (2015:3) that because of the crucial difference between English instruction at school and the expectations regarding performance in academic English at university, one cannot solely rely on school marks to identify at risk students. When one looks at the performance statistics of students on a number of indicators, the better predictor, they found, was how students fared on an academic literacy intervention of longer duration. How students were identified, in their case, as being eligible for such an intervention, derived from the prior administration of an academic literacy test. So while the marks obtained on a longer intervention were more reliable predictors of performance in the first year, an academic literacy test was still used at the beginning to place them on such a course. One could argue that a longer, and hence potentially more reliable assessment, might give one an even better chance of improving the predictive quality of such a test. Moreover, if an appropriate level test is administered early, for example in the final years of secondary school, it may still be a useful indication not only of current, but perhaps also of future performance (for example at tertiary level, or in the world of work).

The temptation to use the National Benchmark Tests (NBTs) as access tests derives in part from them being administered before entry to university. These tests were designed "to better inform learners and universities about the level of academic support that may be required for successful completion of programmes" (National Benchmark Tests Project 2013), which clearly categorises the NBTs as placement tests. Yet because they are written before university enrolment, some universities and tertiary educational institutions use the results of the NBTs to accept or deny students access to their programmes. This is not entirely defensible, as it contradicts the purpose of the test, which is that of a placement test. Cliff and Hanslo (2005:1) note

that it "goes almost without saying that Higher Education institutions worldwide, and the coordinators of the study programmes these institutions offer, need to adopt a coherent and defensible approach towards the selection of students to these institutions". Selection can only refer to an access decision, while placement on a language intervention after entry has been granted is not a determinant of being allowed in, but rather a lower stakes diagnosis of what kind of language development intervention is required and appropriate. The first kind of decision is a high stakes decision that will have effects on the increased or limited earning power of an individual student throughout their working lives. The latter kind is a medium to low stakes decision about what kind of post-admission support might be appropriate for students to develop their ability to handle academic discourse at university. The defensibility of using the academic and quantitative literacy ('AQL') component of the NBT for predicting performance has been questioned in a study that was recently undertaken on students of a university of technology; as Sebolai's (2016) analyses indicate, the predictive validity of that test for future performance is not only suspect, but non-existent, even as an incremental indicator together with others (for other discussions of the situation at different universities, and a possibly more nuanced interpretation, see Fleisch, Schöer & Cliff 2015; Van Rooy and Van Rooy-Coetzee 2015; also Scholtz 2015). At the higher education institution where Sebolai's (2016) study was done, the only (incrementally) better predictor among the academic literacy and other tests of language ability employed in this higher education context is the Test of Academic Literacy Levels (TALL), which is, perhaps not so incidentally, also the most thoroughly scrutinized test in the assessment literature (see the more than 70 analyses, in the form of doctoral theses, master's dissertations and scholarly publications in accredited journals and books that are listed on the 'Research' tab of the ICELDA website: ICELDA 2017; for examples of where the NBT has been scrutinized, see Cliff 2014). It should be noted, in addition, that Sebolai's (2016) study focussed not only on the use of TALL and the NBT, but on all of the various (and in cases highly problematic) language assessments in use at his institution. So, while these findings may perhaps not be generalizable to other environments, this paper takes as its starting point the latter, apparently more appropriate, kind of academic literacy test.

The more desirable eventuality, as Van Rooy and Van Rooy-Coetzee (2015) indicate, is that an academic literacy test taken at an even earlier stage is needed. Such a test might indicate the level of academic literacy of a prospective student at a much earlier time, as well as what kind of academic literacy instruction should be provided in order to prepare that student better for eventually being able to handle academic discourse at university level. If such assessment of the ability to handle academic discourse is administered earlier rather than on or directly before arrival at university, it might by implication also be beneficial to upper secondary school students and

their teachers, by raising the kind of awareness that the curriculum indeed already requires: that they should be able to meet the demands of academic discourse beyond school. Such a test must, however, be theoretically defensible, a point which we shall first discuss below.

3. Designing theoretically defensible assessments of language

Weideman (2011) identifies at least three key principles worth following in designing language assessments. Firstly, test designers should articulate a test construct which outlines the purpose and character of the desired test. The construct defines this purpose and supports the construct validity of a test, or what Weir (2005) calls the theory-based validity of a test. In the case of the tests relevant to this study, we should note that language is dependent on the educational and academic context in which it is presented (Patterson & Weideman 2013:109). For a language assessment to have contextual relevance implies that a variety of specific functional language acts might need to be articulated for the typically different language context it is intended for. Such variable contexts of use have been defined as language use in a variety of material lingual spheres, or discourse types. One would therefore need to establish what combination of language acts is needed for a student to function competently in an academic context. Patterson and Weideman (2013:118) propose the following definition of academic discourse:

> Academic discourse… includes all lingual activities associated with academia, the output of research being perhaps the most important. The typicality of academic discourse is derived from the unique distinction-making activity which is associated with the analytical or logical mode of experience.

From the definition given above two things can be inferred: firstly that the analytically stamped act of distinction-making is central to one's interaction with academic texts, and secondly, by implication, that other complementary acts may also be identified. Once these acts, making up the various components of academic discourse, have been identified, one can proceed to design a test consisting of various tasks and test items that measure the said components. The functionally defined components of academic literacy being referred to here constitute the construct of a test that measures this ability. Such a construct is directly linked to one's idea of academic literacy and what level of ability can be expected of students for them to be able to handle the demands of academic discourse. According to Blanton's definition of academic literacy (1994:226), for example, students should be able to:

1. interpret texts in light of their own experience and their own experience in light of texts;
2. agree or disagree with texts in light of experience;

3. link texts to each other;
4. synthesize texts, and use their synthesis to build new assertions;
5. extrapolate from texts;
6. create their own texts, doing any of the above;
7. talk and write about doing any or all of the above;
8. do number 6 and 7 in such a way to meet the expectations of their audience.

Although an enlightening list, it does not include detail of some of the subskills also needed by students when engaging with academic texts. Working from earlier definitions of academic literacy, Weideman, Patterson and Pot (2016:7) articulate a more extensive list of skills, or as they term it, components of academic literacy, which include the ability of students to:

- ✓ understand a range of academic vocabulary in context;
- ✓ interpret and use metaphor and idiom, and perceive connotation, word play and ambiguity;
- ✓ understand relations between different parts of a text, be aware of the logical development of (an academic) text, via introductions to conclusions, and know how to use language that serves to make the difference parts of a text hang together;
- ✓ interpret different kinds of text type (genre), and show sensitivity for the meaning that they convey, and the audience that they are aimed at;
- ✓ interpret, use and produce information presented in graphic or visual format;
- ✓ make distinctions between essential and non-essential information, fact and opinion, propositions and arguments; distinguish between the cause and effect, classify, categorise and handle data that make comparisons;
- ✓ see sequence and order, do simple numerical estimations and computations that are relevant to academic information, that allow comparisons to be made, and can be applied for purposes of an argument;
- ✓ know what counts as evidence for an argument, extrapolate from information by making inferences, and apply the information or its implications to other cases than the one at hand;
- ✓ understand the communicative function of various ways of expression in academic language (such as defining, providing examples, arguing); and
- ✓ make meaning (e.g. of an academic text) beyond the level of the sentence.

There is a design challenge in transforming these components into a range of task types or subtests that will allow an assessment of the level of mastery of the components of academic literacy articulated above. Below is a table of these components or test specifications and the task types that potentially align with them (Van Dyk and Weideman 2004:18-19):

Specification/ component	Possible task types
Vocabulary comprehension	Vocabulary knowledge Dictionary definitions Cloze C-procedure
Understanding metaphor and idiom	Longer reading passages
Textuality (cohesion and grammar)	Scrambled text Cloze C-procedure (perhaps) Register and text type Longer reading passages Academic writing tasks
Understanding text type (genre)	Register and text type Interpreting and understanding visual & graphic information Scrambled text Cloze procedure Longer reading passages Academic writing tasks (possibly also) C-procedure
Understanding visual & graphic information	Interpreting and understanding visual & graphic information (potentially) Longer reading passages
Distinguishing between essential/non-essential information	Longer reading passages Interpreting and understanding visual & graphic information Academic writing tasks
Numerical computation	Interpreting and understanding visual and graphic information Longer reading passages
Extrapolation and application; finding evidence for an argument	Longer reading passages Academic writing tasks (Interpreting and understanding visual & graphic information)
Communicative function	Longer reading passages (possibly also) Cloze, scrambled text
Making meaning beyond the sentence	Longer reading passages Register and text type Scrambled text Interpreting and understanding visual & graphic information

Table 1: Test components and specifications

What is useful about this list is that each task type can measure more than one component at a time. Therefore, by choosing a certain selection of task types, one can design a practical test which does not have to take several hours to write and can,

therefore, be administered more easily. Since all the tasks are in multiple-choice format, it is a challenge to make items that test a functionally defined construct of academic literacy, in contrast to the kinds of items in a skills-based construct. For example, we may decide to test whether there is genre-sensitivity, an understanding of text type, by asking candidates to match some sentences with sentences from similar texts, as in the following example from the theme-based test on music in a book of practice tests (Weideman & Van Dyk 2014):

> **The sentences below are examples of different text types, such as advertisements, interviews, academic textbooks and the like. You must match an item from the first set (51-35) with an item from the second set (A-E)**
>
> 1. The Beatles were an English rock band, and one of the most commercially successful acts in the history of popular music.
>
> 2. MTV Games and all related titles and logos are trademarks of MTV Networks, a division of Viacom International Inc.
>
> 3. Unlike most hollow-bodied Rickenbackers, it appears to be a solid-body until one picks it up and feels the unusually light weight.
>
> A. ©2009 Harmonix Music Systems, Inc. All rights reserved. The Beatles: Rock Band developed by Harmonix Music Systems, Inc.
>
> B. He took out the original whammy bar and replaced it with the Bigsby vibrato pedal, and, in 1962, he gave it a black finish.
>
> C. According to RIAA certifications, they have sold more albums in the United States than any other artist.

Or one may wish to test whether a candidate is able to find evidence for an argument, or making meaning beyond the sentence by using questions such as the following in a text comprehension task (also taken from the Music test in Weideman & Van Dyk 2014):

> 59. Evidence for the answer to the previous question can be found in the phrase
>
> A. a new book about the origins of music in the delta.
> B. “the blues had a baby … they named rock and roll.”
> C. he traces the blues, a seminal influence ... back to its roots.
> D. whose very privation inspired an impassioned … culture.
>
> 67. A description for the unique sound of the delta-blue music style is given in paragraphs
>
> A. 1 and 2
> B. 2 and 3
> C. 3 and 4
> D. 4 and 5

4. Further principles of responsible test design

The articulation of the construct, as outlined in the previous section, together with its operationalisation in various task types and test items, as in Table 1 above, is done in order to satisfy the conventionally agreed principle of assessment design that relates to its theoretical defensibility. Various further factors must, however, be considered when designing an academic literacy test.

One such requirement is that the texts used in such a test should be at the appropriate level for the intended group of test takers, another that the test should be reliably scored, and yet another that the results should be useful and have credibility. Appropriateness, reliability and practicality are therefore three further important conditions for responsible test design. Regarding appropriateness, the texts used for a test aimed at a specific set of students should be graded on a relevant level for those students. For Grade 10 students, for example, the Flesch reading ease score of a text should preferably be above 50% and fall within a Grade 10 level of difficulty (Steyn 2010:5).

To design an assessment which has to measure the academic literacy abilities of Grade 10 students in such a way that it can be reliably scored adds several further considerations. The test needs to be technically consistent as can be measured using a reliability index such as Cronbach alpha or Greatest Lower Bound (Weideman 2011:105). When a test measures consistently, it will generate similar results when administered to the same group of students on different occasions. Overall test consistency is dependent on the performance of subtests, and, eventually, on how productive individual items in the test are. Test items which do not perform well when tested can be replaced or refined using indices such as Cronbach alpha at test level, combined with measures of discriminatory ability at item level.

All of these factors are principles of responsible test design that are related to the insight that a language test is qualified by its technical function of design (Weideman 2014). For example, since the leading or qualifying technical modality of this applied linguistic artefact has a reciprocal relationship with its analytical dimension, the grounding of the design in current theory in order to provide it with a rationale is a principle of test design; hence the discussion above about the construct of the test, and the further operationalisation of that construct in a set of specifications relating to task (subtest) and item type. The leading technical aspect of a test therefore guides the design of a test, while the analytical dimension generates the founding theoretical rationale behind the design (Du Plessis 2012:36). In a similar way, the technical reliability or consistency of a test referred to above is dependent on the relationship that exists between the technical mode of experience and the kinematic dimension of reality. Each connection of the leading technical function of a test with other

dimensions of experience yields another normative design condition or principle. In all, Weideman (2014:8) distinguishes 14 such design principles:

- ✓ Systematically integrate multiple sets of evidence in arguing for validity of the test or course design.
- ✓ Specify clearly and to the users of the design, and where possible to the public, the appropriately limited scope of the instrument or the intervention, and exercise humility in doing so.
- ✓ Ensure that the measurements obtained and the instructional opportunities envisaged are adequately consistent.
- ✓ Ensure effective measurement or instruction by using defensibly adequate instruments or material.
- ✓ Have an appropriately and adequately differentiated course or test.
- ✓ Make the course or the test intuitively appealing and acceptable.
- ✓ Mount a theoretical defence of what is taught and tested in the most current terms.
- ✓ Make sure that the test yields interpretable and meaningful results, and that the course is intelligible and clear in all respects.
- ✓ Make not only the course or the test, but information about them, accessible to as many as are affected by them.
- ✓ Present the course and obtain the test results efficiently and ensure that both are useful.
- ✓ Mutually align the test with the instruction that will either follow or precede it, and both test and instruction as closely as possible with the learning.
- ✓ Be prepared to give an account to the users as well as to the public of how the test has been used, or what the course is likely to accomplish.
- ✓ Value the integrity of the test and the course; make no compromises of quality that will undermine their status as instruments that are fair to everyone, and that have been designed with care and love.
- ✓ Spare no effort to make the course and the test appropriately trustworthy and reputable.

The analogical moments and other dimensions of reality that are reflected in the technical can each be taken up as an injunction to language test designers to create tests that conform to certain fundamental principles. When a test conforms to these, that provides a greater likelihood for the test and its construct to be theoretically defensible, or for the assessment as a whole to have been responsibly designed. In short, what is called "responsible design" in this framework, is what is usually identified as the factors that contribute to a more successful 'validation' argument for what is conventionally termed the validity of a test.

A key principle among those articulated above relates to the appropriate interpretation of test results. Simply having a pass or fail option for measurements as complex as academic literacy tests does not suffice and leads to the possible inappropriate stigmatization of students, e.g. as being either clever or not. By using a risk band system instead to classify performance and make sense of results, students' abilities are arranged along a spectrum of possibilities which indicates a

student's level of risk as regards language ability. Such a system is at the same time not only more informative, but also more useful and humane than simply having students pass or fail. Following the principles of responsible test design does not ensure that a test is faultless, but it can assist in bringing into harmony the intention and design of a test with its results.

Designing a test as meticulously and deliberately as in the current case also implies that one has to be strictly mindful of its construct and purpose, as has been argued above. This awareness includes keeping in mind all the time what the intended target audience is, since test items and content are methodically modelled after the needs of the identified target group. To be a measurement that is appropriate for the social context in which it will be employed is indeed a principle of responsible test design.

5. Target population

A total of 242 Grade 10 students was the target group for this study. The most significant reason for their selection was their grade level, on the assumption that the early identification of at risk students in need of academic literacy support will be beneficial. Identifying students who struggle to engage with academic texts at Grade 10 level would offer schools and parents more time to prepare students for the academic demands of tertiary educational institutions. This also implies that the curriculum should be more precise than simply stating that students should be "able to use a sufficiently high standard of language in order to be able to gain access to further or Higher Education" (Department of Basic Education 2011:9).

The test was administered to two separate groups of Grade 10 students in the Bloemfontein area in central South Africa. The one group (n=162) forms part of a school formerly identified as a Model C school which is well-known for being well-resourced, and for its academic performance. The second group (n=80) may be labelled as a township school with more limited funds and resources. The school is in the more privileged, less disadvantaged formerly 'coloured' section of the township. By selecting schools that are respectively, in broad terms, well-resourced and potentially under-resourced, it was assumed that differences (and unfair discrimination based on this selection, as would be shown in Differential Item Functioning [DIF] analyses) could be investigated. It should also be mentioned, however, that the analysis might show that the differences might not eventually be as great as would have been the case if, say, the first school, with its reputation of being a top performing school, were compared with an entirely dysfunctional school in a desperately poor area, rather than with a moderately well to do one (by South African standards) in a less disadvantaged part of an urban township.

6. *"*Gadgets and freaky inventions*"*: motivation for the test selected

The main test used in this study, *Gadgets and freaky inventions*, was taken from a book of practice tests compiled by Weideman and Van Dyk (2014). The test was considered most appropriate for the target group, and is a theme-based assessment on *Gadgets and freaky inventions.* Not only was the theme of "gadgets and freaky inventions" considered to be highly relevant for the technologically savvy test population, but the texts selected were also, according to the measures that will be discussed below, deemed to be at the right level of difficulty and aligned with the grade level of the students. This test was not only designed according to the definition of academic literacy outlined above, but an alignment is apparent amongst the test construct, test components and task specifications (Myburgh 2015: 59). The original test total of 100 marks was reduced to 60 marks using the test specifications from another study (Steyn 2015) in which the academic literacy levels of Grade 12 students were tested. With the help of a high school teacher, a few questions in the test were omitted or further adapted for the target group of Grade 10 students. The list of specifications for subtests from the Grade 12 study can be seen below:

Subtest and general task type	**Component (potentially) measured**	**Specifications for items (60 marks)**
A "**Scrambled text**" in which the candidate is given an altered sequence of sentences and must determine the correct order in which these sentences must be placed.	**T**extuality: cohesion and grammar, understand relations between different parts of a text **S**ee sequence and order **U**nderstanding text type (genre) **C**ommunicative function **M**aking meaning beyond the sentence	(5) ✓ Sequencing
"**Vocabulary knowledge**" is tested in the form of multiple choice questions	**V**ocabulary comprehension: understand and use a range of academic vocabulary (limited to a single sentence)	(10) ✓ Vocabulary in context (use) ✓ Handling metaphor and idiom (optional)
The "**Interpreting graphs and visual information**" subtest consists of questions on graphs and simple numerical computations.	**U**nderstanding text type (genre) **U**nderstanding graphic and visual information **D**istinguish between essential and non-essential information **N**umerical computation	(8) ✓ Trends: ✓ Proportions: ✓ Differences between categories ✓ Comparisons of categories

	Extrapolation and application **M**aking meaning beyond the sentence	✓ Inferencing/extrapolation based on the given graphic information.
In the "**Text comprehension**" section, candidates must answer questions about the given text.	**V**ocabulary comprehension **U**nderstanding metaphor and idiom and vocabulary in use **D**istinguish between essential and non-essential information **E**xtrapolation and application **T**hink critically and reason logically and systematically **I**nteract with texts: analyse, link texts, draw logical conclusions **S**ynthesise and integrate information **C**ommunicative function **M**aking meaning beyond the sentence **T**extuality (cohesion and grammar) **U**nderstanding text type (genre)	(25) Essential: ✓ Distinction making (5) ✓ Inferencing/extrapolation (3) ✓ Comparing text with text (2) ✓ Vocabulary in context (5) ✓ Handling metaphor, idiom and word play (1) Another (4) from any of these. Possible: (5) of the following: ✓ Communicative function: e.g. defining/concluding ✓ Cohesion/cohesive ties ✓ Sequencing/text organisation and structure ✓ Calculation
In the "**Grammar and text relations**" section the questions require the candidate to determine where words may have been deleted and which words belong in certain places	**V**ocabulary comprehension **T**extuality (cohesion and grammar) **U**nderstanding text type (genre) **C**ommunicative function	(12) The text is systematically mutilated – a range of components are likely to be measured.

Table 2: Test specifications

From the table one can identify the five subtests as Scrambled text; Vocabulary knowledge; Understanding graphs and visual information; Text comprehension; and Grammar and text relations. The subtests each measure more than one of the components pertaining to academic literacy (in the middle column; see too Table 1). Consequently, each one of the identified components of academic literacy is then potentially measured by more than one subtest of the same test. Textuality, for example, can be measured by means of a subtest such as Scrambled text, Text comprehension or Grammar and text relations, or all of them.

In order to develop an assessment that test takers would be able to complete more quickly, the original 100 mark version of the main test was modified to a 60 mark

test. The Scrambled text subtest was kept exactly the same; the original also constituted five marks. The remaining subtests were all modified in light of the specifications listed above. Questions were chosen with the assistance of the teacher mentioned and those which were more likely to be misinterpreted by students were discarded. The Verbal reasoning subtest was eliminated altogether, as was the Register and text type subtest. Text comprehension had to be modified to constitute 25 marks instead of 35. Lastly, for Grammar and text relations some of the original questions were retained as examples, whilst the remaining questions were kept as they were.

In addition to the modification of test items and subtests, the texts used for the main test were also analysed to ensure that they were appropriate for Grade 10 students. The Flesch reading ease of a text for Grade 10 students should preferably be above 50% and should fall within a Grade 10 level. This would indicate that the text is neither too difficult nor too easy to read (Steyn 2010:5). The first text within the test has a Flesch reading ease of 56.3% and a Flesch-Kincaid level of 10.5, whilst the second text has a Flesch reading ease of 67% and a Flesch-Kincaid level of 8.6 (Steyn 2010:5).

As part of the experiment, a second test of academic literacy, the slightly higher level (Grade 12) Test of Advanced Language Ability (TALA) (Steyn 2010, 2015), was also administered to students, though for the sake of brevity its results are given below only to provide comparative data.

7. Method

Three comparisons were carried out on the data captured. The first comparison was between the results obtained in the test and the students' Home Language mark. Secondly, the test result was also compared to the students' average mark across all subjects. Lastly, the results were once again compared to the students' average mark, but this time their Home Language mark was omitted from their overall average mark. The aim was to determine whether the main assessment, the *Gadgets and freaky inventions* test, would more accurately predict the students' average mark than the Home Language mark would. Additional analyses were carried out on the data to determine test and item performance. These included an Iteman 3.6 and Iteman 4.3 analysis, as well as a TiaPlus analysis.

8. Results

An Iteman 3.6 analysis (Assessment Systems Corporation 2006) indicated that the main test scored a Cronbach alpha of 0.896, which is well above the required 0.7

score that academic literacy tests in this context usually aim for. Iteman 4.3, which is a more recent version of the program, and which provides additional statistics and information regarding a test (Guyer & Thompson 2011), indicated an alpha score of 0.897 for *Gadgets and freaky inventions* in this administration.

Another statistical analysis, done with TiaPlus, measured the intercorrelations between subtests, which gives one a partial indication of the construct validity of the test, or lack thereof (Du Plessis 2012:130). On what may be considered a conservative set of parameters, subtest intercorrelations should fall between 0.3 and 0.5 (Van der Walt & Steyn 2007), since one is seeking neither too close, nor too distant a correlation among components of a test measuring the same ability. In addition, one is looking for a higher correlation (of above 0.7) between the subtest and the test as a whole. The values for this test and its component subtests are presented in the table below:

Subtest		Test	1	2	3	4	5
Scrambled text	1	0.64					
Vocabulary know	2	**0.70**	**0.42**				
Interpreting gr	3	**0.81**	**0.50**	**0.49**			
Text comprehens	4	**0.89**	**0.43**	0.57	0.64		
Grammar & text	5	**0.74**	**0.33**	**0.37**	**0.52**	**0.54**	
Number of testees	:	240	240	240	240	240	240
Number of items	:	60	5	10	8	25	12
Average test score	:	33.23	2.51	6.23	5.00	13.65	5.83
Standard deviation	:	10.53	1.99	1.75	2.43	4.44	2.83
SEM	:	3.40	0.74	1.30	1.09	2.22	1.54
Average P-value	:	**55.39**	50.25	62.33	62.55	54.62	48.58
Coefficient Alpha	:	**0.90**	0.86	0.44	0.80	0.75	0.70
GLB	:	**0.97**	0.90	0.64	0.85	0.86	0.89
Asymptotic GLB	:	0.96	0.90	0.53	0.84	0.84	0.84

Table 3: Subtest intercorrelations of the second test

Of the ten subtest intercorrelations, eight fall within the preferred parameters, whilst only two subtest intercorrelations can be regarded as possibly too strong (0.57 and 0.64). At the same time, four of the five correlations between the subtests and the test as a whole fall within the specified parameters, whilst only one correlation is slightly too low.

Another useful statistic given by TiaPlus is Differential Item Functioning (CITO 2005). DIF indicates whether items within a test are potentially biased towards certain groups. This is important for this study since the two groups which were used for this study were assumed to be socio-economically divergent. Concerning DIF, TiaPlus indicated that there were no items which were biased towards one of the groups, which is wholly satisfactory given the circumstances of the test administration. Whether the same degree of lack of DIF would be evident if the

results from a really under-resourced, dysfunctional school were included in the sample and analysis, is of course another question. The only answer one can currently give to such a hypothetical case is that one does not know.

Various further data analyses were executed on the data by the Statistical Consultation Unit (SCU) at the University of the Free State, including a regression analysis, a correlational analysis and an ANCOVA analysis (Statistical Consultation Unit 2014; discussed in detail in Myburgh's 2015 study). A regression analysis was completed on the results acquired through the administration of three measurement devices, including the two academic literacy tests already mentioned, TALA and *Gadgets and freaky inventions*, and the English Home Language school examination paper of June 2014 on two Bloemfontein based schools. The aim of the analyses is to establish whether notable comparisons exist between the academic performance of the students over all their various subjects (usually referred to as a student's average) and the results the students obtained for the three said measurement devices. The results of specifically the correlational analysis (extracted from Annexure H of Myburgh 2015:269) can be seen in the table given below.

	Average without English (p)	**Test 2 [Gadgets…] (p)**	**Test 3 [English] (p)**
Average without English	1.00000	0.78491 (<.0001)	0.81810 (<.0001)
Test 1 [TALA] (p)	0.45512 (<.0001)	0.35253 (<.0001)	0.31814 (<.0001)
Test 2 [Gadgets…] (p)	0.78491 (<.0001)	1.00000	0.78408 (<.0001)
Test 3 [English] (p)	0.81810 (<.0001)	0.78408 (<.0001)	1.00000

Table 4: Correlational analysis results

From Table 4 it can be seen that the students' Home Language mark [English] predicted more accurately the students' average mark with a correlation of 0.81810, whilst the *Gadgets and freaky inventions* test (Test 2 [Gadgets…]) predicted the students' average mark slightly less accurately, with a score of 0.78491. Whilst it was disappointing that the *Gadgets and freaky inventions* test did not predict the students' academic performance more accurately than the Home Language mark, a few comments will be made below, in the next section.

9. Discussion of results

It should be noted that a test which was developed by test designers and adapted by a student and teacher for Grade 10 students, and which was administered to students for the first time during this study in 2014, predicted academic performance almost as accurately as 10 preceding years of accumulated assessments and training done by teachers in the South African schooling system. Not only are students prepared in advance for school tests and examinations, which was not the case with the administration of *Gadgets and freaky inventions*, but students complete tasks and homework assignments on a regular basis as well. It should, therefore, be considered noteworthy that a test that learners were unprepared for can nonetheless predict academic performance almost as accurately as their conventional assessments. Given that this test can still be further refined, modified and re-piloted, as will be discussed in the last part of this article, it might well be able to predict better still.

Of course there are limitations to such endeavours and analyses that would still need further probing. The assumption of a degree of fit between average academic performance and the results of an academic literacy assessment does not give us the whole picture, and itself needs further exploration and analysis, perhaps as was done by Van Rooy and Coetzee-Van Rooy (2015). It is beyond the scope of this article to go into all of these potential limitations, but that does not invalidate the claim that they should in future investigations figure prominently.

It is, however, still worth asking: if an academic literacy test can then predict almost as accurately as 10 years of preceding teaching and accumulative assessment, should it not be regarded as an additional option of assessing students at an earlier stage than Grade 12? Moreover, since the test has been designed so diligently in accordance with its construct, detailed test results in the form of feedback reports can be given to indicate the components of academic literacy with which students struggled, or even which ones they excelled in. The test can, in other words, conceivably be employed to yield specific and highly relevant diagnostic information. Ultimately, students would then be able to prepare more appropriately for the demands that tertiary education institutions pose in terms of academic language ability.

10. The refinement of "Gadgets and freaky inventions"

The refinement of a test includes the modification of test items which did not perform as desirably as they should have in light of the Iteman and TiaPlus analyses mentioned previously. The test under consideration here is worthy of refinement also since it came close to predicting academic performance as well as the English Home Language marks did.

There are several parameters of item productivity for the test used in this study. First, the Rpbis score of a correct item should be higher than any of the other incorrect options given for that same item. The Rpbis score, that is a measure of the ability of the item to discriminate among test takers of low and high ability, should be a positive number and should preferably be above 0.15. Second, the P-value of an item should be in the vicinity of 0.5 (Guyer & Thompson 2011), but for this study we have chosen to accept values ranging from 0.2 to 0.8 as suitable. The relevant values (Rpbis and P- or facility value) for the 10 items of the second test which did not perform within these parameters are listed below:

	Rpbis	**P-value**
Item 6	**-0.434**	0.596
Item 7	0.235	**0.904**
Item 12	0.320	**0.929**
Item 13	**-0.118**	**0.129**
Item 25	**0.111**	0.517
Item 27	**0.135**	0.658
Item 28	**-0.045**	**0.179**
Item 32	**-0.129**	**0.146**
Item 45	**0.114**	0.429
Item 52	**0.091**	0.383

Table 5: Summary of items which did not perform satisfactorily, as indicated by Iteman 4.3

Items can simply be removed from a test, which reduces the number of items in a test if they are not subsequently replaced by others, for example, by items that have performed well in other pilots. On the other hand, one may keep to the possible refinement of the items mentioned above based on information taken from the Iteman 4.3 analysis.

In line with that, the wording of Item 6 was examined again, leading to the conclusion that it was most likely construed as ambiguous by test takers and was therefore changed. Most test takers answered Item 7 correctly, indicating that the item might have been too easy. In this instance, the possible answers were changed, in an attempt to make it more difficult. The same pattern was evident for Item 12 and the possible answers were also changed. On the other hand, Item 13 was too difficult for the test takers, according to the Iteman 4.3 analysis. Here, once again, the possible answers were made less ambiguous. For Item 25 it seems that many successful test takers chose the incorrect answer. Therefore, the incorrect answer which was so often chosen was modified so the actual answer would be a clearer choice. The same

occurrence transpired for Item 27, which means more emphasis had to be placed on the correct answer. The same pattern was evident for Item 28, leading to a change in the wording. Item 32 might have been too difficult for the test takers. This was remedied by changing the order of the possible answers. The phrasing of item 45 seems flawed and was therefore modified. Lastly, the possible answers for item 52 were indicated as being problematic, and were thus edited.

The refined version should of course preferably once again be administered to a group of test takers, and can possibly be put through the above mentioned analyses once more. The refined test should then be an improved version of the *Gadgets and freaky inventions* test, and could well in its refined format predict the academic success of the test takers even more accurately than its predecessor (Myburgh 2015:105-109).

11. Conclusion

This paper is intended as a further broadening of the ongoing discussions about the preparedness of secondary school learners to cope with the language demands they will face in higher education.

It takes further the conclusion reached by Sebolai's (2016) recent study that a well-designed, deliberately constructed and theoretically defensible assessment of academic literacy of the kind used in the experiment reported on here can contribute incrementally to our insight into the relation between language ability and successful further study. The points made in other discussions and analyses that were referred to above are equally valid: that longer term interventions are good indications of performance in higher education contexts. But the further point must be that one should have a means, a reliable and useful assessment, of who needs such interventions most, in order to place candidates on them at the earliest opportunity. In the case of this study, Grade 10 was chosen as such an early point of identification, but other work, for example by Grühn (2015) and Steyn (2014), indicates that it can happen a great deal earlier still. The reference to these studies of emergent and early literacy brings us to emphasise the final point once again: the importance that is placed in the curriculum on being prepared for further study as regards one's level of language ability is largely being ignored or neglected (Du Plessis 2017). It is our hope that this contribution to the discussion will help to raise awareness of that neglect, and that it will eventually serve to assist in rectifying it.

Greater awareness of and attention to the ability to handle academic discourse, as required by CAPS, needs to be reflected not only in assessment, but also in language instruction at school. A good place to start would be to raise awareness of designing

language assessments at that level that are theoretically more defensible than, for example, the currently contested home language examinations (as concluded by Du Plessis 2017). At the same time, they should also be much more sophisticated, refined and deliberate.

References

Assessment Systems Corporation. 2006. *User's manual for Iteman 3.6 conventional item analysis program*. St Paul, Minnesota: Assessment Systems Corporation.

CITO. 2005. *TiaPlus: Classical test and item analysis*. Arnhem: Cito Measurement and Research Department.

Cliff, A. 2014. Entry-level students' reading abilities and what these abilities might mean for academic readiness. *Language Matters*, 45 (3), 313 - 324.

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.

Cliff, A. & Hanslo, M. 2005. The use of 'alternate' assessments as contributors to processes for selecting applicants to Health Sciences faculties. *Southern African Linguistics and Applied Language Studies* 24(1):1-19. DOI: http://dx.doi.org/10.2989/SALALS.2009.27.3.5.939.

Cliff, A.F., Yeld, N., Hanslo, M. 2003. Assessing the academic literacy skills of entry-level students, using the Placement Test in English for Educational Purposes (PTEEP). *Bi-annual conference of the European Association for Research in Learning and Instruction (EARLI)*, Padova, Italy, August 2003.

Department of Basic Education. 2001. *National plan for higher education in South Africa*. Pretoria: Department of Basic Education.

Department of Basic Education. 2005. *Student enrolment planning in public higher education*. Pretoria: Department of Basic Education.

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

Du Plessis, C. 2012. *The design, refinement and reception of a test of academic literacy for postgraduate students*. MA dissertation. University of the Free State. URI: http://hdl.handle.net/11660/1353.

Du Plessis, C. 2014a. Issues of validity and generalisability in the Grade 12 English Home Language examination. *Per Linguam* 30(2):1-19. DOI: http://dx.doi.org/10.5785/30-2-602.

Du Plessis, C. 2014b. Writing as construct in the Grade 12 Home Language curriculum and examination. *Journal for Language Teaching* 48(2):121-141. DOI: http://dx.doi.org/10.4314/jlt.v48i42.6.

Du Plessis, C. 2017. Developing a theoretical rationale for the attainment of greater equivalence of standard in the Grade 12 Home Language exit-level examinations. PhD thesis. University of the Free State. (Unpublished).

Du Plessis, C., Steyn S. & Weideman, A. 2014. Towards a construct for assessing high level language ability in the South African National Certificate. Submitted to *Africa Education Review*.

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

Grühn, S. 2015. Initial validation of a test of emergent literacy. MA dissertation. Rijksuniversiteit, Groningen. Available: http://arts.studenttheses.ub.rug.nl/16752/ [Accessed: 17 March 2017.]

Guyer, R. & Thompson, N. 2011. *User's manual for Iteman 4.2*. St Paul, Minnesota: Assessment Systems Corporation.

Inter-Institutional Centre for Language Development and Assessment (ICELDA). 2017. Research. Available: http://icelda.sun.ac.za. [Accessed: 17 March 2017.]

Jeffery, A. 2014. Goeie en slegte nuus van 20 jaar. *Rapport*. 25 April, p.7.

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. MA dissertation. University of the Free State. URI: http://hdl.handle.net/11660/2081.

National Benchmark Tests Project. 2013. Placement testing. Available: http://www.nbt.ac.za/content/placement-testing. [Accessed: 16 March 2017].

Patterson, R. & Weideman, A. 2013. The typicality of academic discourse and its relevance for constructs of academic literacy. *Journal for Language Teaching* 47(1):107-123. DOI: http://dx.doi.org/10.4314/jlt.v47il.5.

Scholtz, D. 2015. *A comparative analysis of academic literacy specifications for a standardised test and academic literacy requirements for reading and writing in a range of disciplinary contexts*. DPhil thesis, University of Cape Town. Available: http://hdl.handle.net/11427/16866

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. URI: http://hdl.handle.net/11660/5408.

Statistical Consultation Unit. 2014. Bloemfontein: University of the Free State.

Statistics South Africa. 2017. Educational enrolment and achievement, 2016. (Education series volume III; Report 92-01-03). Pretoria: Stats SA. Available: http://www.statssa.gov.za/publications/Report%2092-01-03/Report%2092-01-032016.pdf. [Accessed: 17 March 2017.]

Steyn, S. 2010. Devising a test to aid in the process of preparing prospective students for academic language proficiency tests at tertiary institutions. [Unpublished report for ICELDA].

Steyn, S. 2014. *The design and refinement of a test of early academic literacy.* MA dissertation. Rijksuniversiteit, Groningen. Available: http://arts.studenttheses.ub.rug.nl/15198/. [Accessed: 17 March 2017.]

Steyn, S. 2015. A theoretical justification for the design and refinement of a test of advanced language ability for learners at FET level. In preparation: MA dissertation. University of the Free State.

Van der Walt, J. & Steyn, H. 2007. Pragmatic validation of a test of academic literacy at tertiary level. Ensovoort 11(2):138-153. URI: http://hdl.handle.net/10394/2841.

Van Dyk, T. & Weideman, A. 2004a. Switching constructs: on the selection of an appropriate blueprint for academic literacy assessment. *Journal for*

Language Teaching 38(1):1-13. DOI: http://dx.doi.org/10.4314/jlt.v38i1.6024.

Van Dyk, T. & Weideman, A. 2004b. Finding the right measure: from blueprint to specification to item type. *Journal for Language Teaching* 38(1):15-24. DOI: http://dx.doi.org/10.4314/jlt.v38i1.6025.

Van Rensburg, C. & Weideman, A. 2002. Language proficiency: current strategies, future remedies. *Journal for Language Teaching* 36(1):152-164. DOI: http://dx.doi.org/10.4314/jlt.v36i1-2.6010.

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*. DOI: http://dx.doi.org/10.2989/16073614.2015.1012691.

Weideman, A. 2009. *Beyond expression: a systematic study of the foundations of linguistics*. Grand Rapids: The Reformational Publishing Project.

Weideman, A. 2011. Academic literacy tests: design, development, piloting and refinement. *Journal for Language Teaching* 45(2):100-113. DOI: http://dx.doi.org/10.4314/jlt.v45i2.6.

Weideman, A. 2014. Innovation and reciprocity in applied linguistics. *Literator*. DOI: http://dx.doi.org/10.4102/lit.v35il.1074.

Weideman, A., Patterson, R. & Pot, A. 2016. Construct refinement in tests of academic literacy. In: J. Read (Ed.). *Post-admission language assessment in universities: international perspectives.* Cham: Springer. Chapter 9, p. 179-196. DOI: 10.1007/978-3-319-39192-2_9.6.

Weideman, A. & Van Dyk, T. (Eds.). 2014. *Academic literacy: test your competence*. Bloemfontein: Geronimo Distribution.

Weir, C. 2005. *Language testing and validation: an evidence-based approach*. Hampshire: Palgrave-Macmillan.

Albert Weideman is professor of applied language studies and senior research fellow at the University of the Free State. His latest book is *Responsible design in applied linguistics: Theory and practice* (2017, Springer). He focuses on how language assessment relates to a theory of applied linguistics.

e-mail address: albert.weideman@ufs.ac.za

Jo-Mari Myburgh-Smit's dissertation for the MA in Linguistics dealt with the assessment of academic literacy at senior secondary school level. Her studies focus on responsible and appropriate language assessment. A former NRF-DST intern, she is currently interpreting coordinator at the University of the Free State.

e-mail address: myburghj2@ufs.ac.za