Southern African Linguistics and Applied Language Studies 2008, 26(3): 363–378
Printed in South Africa — All rights reserved

SOUTHERN AFRICAN LINGUISTICS
AND APPLIED LANGUAGE STUDIES
ISSN 1607-3614 EISSN 1727-9461
DOI: 10.2989/SALALS.2008.26.3.5.632

Measures of improvement in academic literacy

Frans van der Slik[1] and Albert Weideman[2]
[1]*Department of Linguistics, Radboud University of Nijmegen*
e-mail: f.v.d.slik@let.ru.nl
[2] *Unit for Academic Literacy, University of Pretoria, Pretoria 0002, South Africa*
e-mail: albert.weideman@up.ac.za

Abstract: This article considers the analysis of the results of a re-administration of a test of academic literacy, specifically with a view to determining whether this analysis yields insight into the improvement of levels of academic literacy over time. The article postulates that, if improvement occurs, the level of improvement will be uneven across different categories of ability, and across different levels of academic literacy as measured by various sub-tests. An attempt is made to offer explanations for the kinds of improvement that are evident, as well as to identify factors that may play a role in such improvement, such as the time of being exposed to a compulsory academic literacy development intervention, the mother tongue of the testee, and the initial level of academic literacy.

Introduction

We have commented before (Van der Slik & Weideman, 2005) on the test developers' quest to make continued refinements to their test designs, and the value of various empirical analyses that are available with this end in view. Specifically, we have concluded that different measures yielded by the statistical properties of a test do not conflict with, but in fact complement, present day concerns about transparency and accountability (Weideman, 2006). We refer here to measures such as the various reliability indices (Cronbach's alpha, Greatest Lower Bound – for a detailed treatment of these in relation to the test currently under discussion, see Weideman & Van der Slik, 2008; the TiaPlus manual (CITO, 2005); also see below), and the calculations associated with them. These calculations can be employed, for example, to identify those who may have been unfairly treated by means of a particular administration of a test, and who should thus be offered a second chance to be assessed.

Even though not all empirical analyses are directly accessible to the general public, a first level of accountability for any test design must remain an intrinsic objective of analyses such as the one that we present here, for the purpose of specific scrutinisation by others within the academic community who are concerned about issues of language testing. The current paper therefore once again takes its cue from Shohamy's (2001) exhortation to 'tell the story of a test' as a necessary first step in the process of becoming transparent and, subsequently, accountable as test developers. For an excellent case study and description of the development of one test across three different eras during the course of a century at one tertiary institution, see Davidson and Cho (2001).

For these reasons, then, this paper comprises a report on further analyses that we have conducted on the results of the Test of Academic Literacy Levels (TALL), as its English version is called, and the TAG [Toets van Akademiese Geletterdheidsvlakke], as it is known in Afrikaans. While the present analysis will limit itself to the TALL, information that is common to the two tests will also be included below, where relevant. These tests are administered annually to all new undergraduate students at several South African universities – North-West University's (NW) Potchefstroom and Vanderbijl Park campuses, the University of Pretoria (UP), and the University of Stellenbosch (US). In 2006, first-year students in the Faculty of Medicine at the University of Limpopo (Medunsa campus) were also assessed by means of TALL. The motivation for the test is primarily economical: extensive wastage occurs within the higher education sector as a result of the

underpreparedness of new arrivals within the university. Since students' (lack of) academic literacy is considered to be an important contributing factor to their lack of performance and success at university (Van Rensburg & Weideman, 2002; see also Cliff *et al.*, 2006; Visser & Hanslo, 2005), the assessment thereof is intended to reveal the extent to which the student is at risk.

Method

Population

In January 2005, the academic literacy of all new undergraduate students of the UP was tested. In total, 2 701 such students participated in the first administration of the Afrikaans test, while 3 310 students sat for the first administration of the English test. In terms of the outcomes of these tests, students' academic literacy could be divided into 5 levels, as described below, from *Extremely high risk* (Level 1) to *Low to no risk* (Level 5). Students who were identified by the test as being at risk as a result of too low a level of academic literacy had to undergo a compulsory language intervention (Weideman, 2003) to improve this ability. Those students at Level 3 (initially identified as *Borderline* cases) who comprise part of this study are those who either did not take the second-chance test offered to borderline cases, or who did undergo the test, but were identified for a second time as being at risk as a result of too low a level of academic literacy, and who were therefore categorised as *At risk*. Like the Level 1 (*Extremely high risk*) and Level 2 (*High risk*) students, they had to participate in a compulsory academic literacy development intervention, which entails a year-long language course. In the English version of the test (TALL) 1 134 such students were identified as being at risk.

In November 2005, many of the students who had been identified in January as being at risk repeated the English version of the academic literacy test. They therefore wrote exactly the same test that they had written in January. Since not all of the students continued with the course, and because the final assessment included a number of students from previous years who still needed either to complete the course as a whole, or to complete part of it, there were fewer students in respect of whom data for both the January and November assessments were available. We do, however, have information on 738 such students.

The test: TALL 2005

The 2005 versions of the TALL/TAG consist of 80 items (TALL) and 82 items (TAG) respectively. These items are distributed over 7 sub-tests or sections (described in Van Dyk & Weideman, 2004a), 6 of which are in multiple choice format:

- Section 1: Scrambled text (5 items)
- Section 2: Knowledge of academic vocabulary (10 items; based on Coxhead 2000; for the theoretical justification in respect thereof, see also Coxhead & Nation, 2001)
- Section 3: Interpreting graphs and visual information (TALL 6 items; TAG 7 items)
- Section 4: Text types (5 items)
- Section 5: Understanding texts (TALL 19 items; TAG 20 items)
- Section 6: Text editing (15 items) (For a discussion of the design of this sub-test, see Van Dyk & Weideman, 2004b)
- Section 7: Writing (handwritten; marked and scored only for certain borderline cases, 20 marks)

The last section mentioned above, writing, had more or less the same format as the International Language Testing System (IELTS) component, 'academic writing,' described and discussed by Coffin and Hewings (2004; see also Davidson & Cho, 2001). In some subsequent versions, this section has been omitted, since the significant correlation between the marks awarded for it and the rest of the test was of such a nature that its results offered little additional information regarding academic literacy, while the scoring process was arduous. Also, it was argued that, in an initial test, it would make more sense to focus on pre-writing skills.

Students were given 60 minutes to complete the test; and they could earn a maximum of 100 points (with approximately half of the items being worth 2 or 3 points, instead of only 1). The results of the test (TALL and TAG) categorise students in terms of their academic literacy level, as follows (Unit for Academic Literacy 2006):

1. *Extremely high risk*
2. *High risk*
3. *Borderline* case (after a second assessment, either identified as *At risk*, or *Low risk*)
4. *Low risk*
5. *Low to no risk*

In Weideman and Van der Slik (2008), we have dealt with the refinement of these categories, and especially with the arguments that might provide a more solid empirical basis for such categorisation. In that article, we have also discussed the different arguments that converge to allow for a justifiable and defensible cut-off point for the administration of the results of this test. Generally, however, the cut-off point for TALL is about 10% below the average score of the particular administration that is involved, and sets the level of High risk at about 30% of the population involved. As regards the determination of borderline (second-chance) cases, the indication is that a responsible cut-off point, which is determined by bringing into play the various measures of reliability (Cronbach's alpha and Greatest Lower Bound; Jackson & Agunwamba, 1977), usually lies between 0.15 and 0.31 standard deviations away from the mean. The arguments are too varied and complex to discuss more fully here, but are dealt with extensively in Weideman and Van der Slik (2008).

As we have remarked above, this paper deals only with the English version of this test (TALL). Once we are satisfied that the analysis yields noteworthy information, we envisage repeating this analysis for the Afrikaans version (TAG). The levels of reliability that we have observed in the various administrations of the English version of the test (TALL), and which therefore comprise relevant background information to the analyses in this paper, are summarised in Table 1.

While reliability measures for the third participating university, the University of Stellenbosch, are lacking for several of its administrations, the average reliability level for the two administrations for which data are available (2005 and 2006), is 0.90. All of this indicates that the test is highly reliable, a point to which we shall return in some of the arguments below.

Background information

In January, when students took the TALL test, they were asked to provide background information about the following:

1. *Faculty* ('A' through 'H'): information on the faculty in which they had enrolled
2. *English as a subject at school* (EngopSkool) ('Yes' or 'No')
3. *Language of instruction at secondary school* (Skooltaal) ('English', 'Afrikaans' or 'Other')
4. *Mother tongue* (Moedertaal) (1 [Afrikaans], 2 [English], 3 [Other South African (African) language] or 4 [Other])
5. *Whether English was studied as a first or additional language at school* ('A' or 'F')
6. *The number of years during which English was studied as a subject* (Engjare)

In addition, information is available regarding the following aspects:

7. *Lecturer* (the lecturer responsible for instruction in the students' academic literacy class; this information is only available for the November results)
8. *Level of academic literacy in January* (1, 2 or 3) (only for January results)

In this study, the criteria, mother tongue and level of academic literacy (1–5 above), have been

Table 1: Reliability levels (in terms of Cronbach's alpha) across several administrations of the Test of Academic Literacy Levels (TALL)

	University of Pretoria	North-West University
2004	0.95	0.92
2005	0.93	0.94
2006	0.94	0.93
2007	0.94	0.94
2008	0.94	0.93
Average	**0.94**	**0.93**

used to predict differences in academic literacy, while 'Lecturer' was used as a control variable. Unfortunately, 25% of the students did not provide valid information about their mother tongue. The reason for this may be that many are simply uncertain as to which language is their mother tongue, since they come from homes and communities in which a variety of languages and the primary use of more than one language are the norm. What is one's first language if one's mother speaks isiXhosa and one's father speaks Setswana, while the aunt who has reared one since childhood is isiZulu-speaking – and the language that one uses to communicate with one's immediate friends is Sesotho? As a result, information in respect of only 610 students has been used in the analyses. In passing, it may be noted that students who provided no valid information on their mother tongue achieved lower scores in the academic literacy test, both in January (t = -3.25, p = 0.001) and November (t = -2.64, p = 0.006). The T-scores for these students indicate that the lower average scores that this group obtained differ significantly from the scores of those who were able to provide valid information on their first language. Because the differences in academic literacy scores between those who provided valid information and those who did not, remain constant between January and November, it is our opinion that this type of non-response will not seriously affect the results discussed below.

Analysis

In order to analyse the test results of the UP undergraduates identified for this study, we made use of the 'repeated-measures analysis of variance'. For the analysis of the total test, we used the test that was taken in January and repeated in November as a 'within-subject' time factor, and level and mother tongue as 'between-subject factors'. The term 'within-subject' is used to describe a factor that concerns, for example, individual students and their progress (or lack thereof) over time. A 'between-subject' factor is one that expresses differences among individuals, for example in respect of differences in mother tongue, lecturer, faculty, school, etc.

For the analysis of the six sub-tests, we used the results of the test taken in January and November as a 'within-subject' Time factor, and each sub-test as a within-subject factor nested in Time. Level of academic literacy and mother tongue were used as 'between-subject' factors.

In doing so, we aimed to determine whether academic literacy shows an improvement over time, not just as measured by the entire test, but also as measured by the six sub-tests constituting the TALL. Our hypothesis was that the analysis would show an improvement in the level of academic literacy over time, but that the improvement might be uneven across levels of ability, and across different kinds of academic literacy tasks – in the latter case, as represented in the different sub-tests.

At the start of the analysis, we first tested to see whether Lecturer had an effect on academic literacy improvement. While we initially found that Lecturer did have a significant effect on test-score improvement ($F_{(14, 595)}$ = 3.11; p < 0.001), and that it therefore seemed as if some lecturers were more successful in improving students' academic literacy than others, this conclusion turned out to be premature and, in fact, completely unwarranted. A non-random dispersion of students with regard to their mother tongue and their initial level of academic literacy, rather than Lecturer, was entirely responsible for the occurrence of this apparent Lecturer effect. In fact, no significant Lecturer effect on test-score improvement remained at all, when the level of academic literacy and mother tongue were accounted for. We therefore omitted the factor, Lecturer, in our eventual analyses.

Results

Table 2 provides the descriptive results in terms of the overall test scores. As can be seen, students' academic literacy seems to improve over time, since their November scores are substantially higher than their January scores. This applies both to Level 1 and to Level 2 students, comprising those who were identified by the initial test as being most in need of improving their academic literacy, and who were therefore required to take compulsory academic literacy modules. The academic literacy of Level 3 students, who were initially classified as Borderline cases and given a second opportunity to have their academic literacy assessed before being re-classified either as At risk or Low risk,

had also improved. In addition, it can be observed that the factor, mother tongue, seems to make a difference. Those students whose mother tongue is Afrikaans or English (mother tongue = 1, 2) achieved substantially higher scores in the TALL than students whose mother tongue is another African language (3) or another language (4), both in January and in November. The reason for this categorisation into Afrikaans and English (1, 2) on the one hand, and other languages (3, 4), on the other, is that Afrikaans and English are the two languages of instruction at the University of Pretoria, and are therefore arguably in a stronger position than any other language (as we shall also see below).

Analyses

Total scores

As already noted, we used 'repeated analyses of variance' to determine which effects are significant. Table 3 gives the outcomes of this analysis regarding the overall test scores.

It can be seen that the within-subject factor Time, which measures the difference between a student's January score and his or her November score, is highly significant ($F_{(1,604)} = 342.25$, $p < 0.001$). The differences in scores that can be seen in Table 2 therefore reveal significantly higher literacy levels in November than the initial levels measured in January. This improvement is, however, dependent on both the initial level of academic literacy and the mother tongue, since the interaction effects of Time with Level and Time with mother tongue are significant ($F_{(2, 604)} = 24.31$, $p < 0.001$, and $F_{(1, 604)} = 5.94$, $p = 0.015$, respectively). The meaning of these findings is further expounded below in Figures 1 and 2, and in the discussion of these figures.

Figure 1 and Figure 2 (below) show how these interaction effects should be interpreted. It seems quite clear that the pace of academic literacy improvement depends on the initial level of academic literacy. Though the comparatively lower-risk students (Level 3) still obtained the highest ratings in November, their skills in academic literacy improved at a slower pace than the academic literacy skills of students who were at greater risk in January (Levels 1 and 2). It seems that the improvement of the Extremely high-risk students (Level 1) is the most substantial, even though, on average,

Table 2: Descriptive statistics of the English version of the academic literacy test in January and November 2005 by level of academic literacy in January 2005 and by mother tongue

		Mother tongue = 1,2			Mother tongue = 3,4			Total		
		Mean	SD	*n*	Mean	SD	n	Mean	SD	*n*
January	Level 1	43.3	8.4	22	38.9	10.1	275	39.2	10.1	297
	Level 2	61.8	4.3	84	60.8	4.7	182	61.1	4.6	266
	Level 3	69.9	.7	17	69.9	0.8	30	69.9	0.8	47
	Total	59.6	9.5	123	49.0	14.2	487	51.1	14.1	610
November	Level 1	66.7	13.7	22	54.7	12.8	275	55.6	13.2	297
	Level 2	74.7	9.1	84	72.5	10.3	182	73.2	10.0	266
	Level 3	77.4	7.4	17	75.0	8.1	30	76.5	7.8	47
	Total	73.6	10.4	123	62.7	14.8	487	64.9	14.7	610

Table 3: ANOVA Repeated measures of the January and November TALL test, with level of academic literacy and mother tongue as between-subjects factors and Time as a within-subject factor

	df	Mean Square	*F*	*p*
Time	1	18 876.23	342.25	0.000
Time * level	2	1 341.03	24.31	0.000
Time * mother tongue	1	327.84	5.94	0.015
Time * level * mother tongue	2	158.81	2.88	0.057
Error (Time)	604	55.15		

they would still fail to make the cut-off point for the TALL (Weideman & Van der Slik 2008). On average, students at risk level 2, who were originally identified as falling into the High risk category, would probably achieve the desired level of academic literacy in November.

There are several ways to interpret these outcomes. Improvement may be owing to:

- a 'testing' effect;
- a 'maturing' effect;
- a 'ceiling' effect;
- a 'regression to the mean' effect; or,
- a 'tutoring' effect.

The possibility of a testing effect may be considered, since the students did write similar tests at the end of the first, second and third terms. Improvement of academic literacy as a result of maturing may also have occurred, as a result of becoming accustomed to the academic environment. The fact that this improvement is greater in the case of the higher-risk students could perhaps also be attributed to a ceiling effect. It is unlikely, however, that this comprises the entire explanation, because a mean score of 76.5 (see Table 2) leaves enough room for further improvement (bearing in mind that the maximum score = 100). A maturing effect might also partly explain the improvement of the higher-risk students. In that case, extra tutoring would perhaps be of minor importance.

To take the matter even further: the extra improvement of the Level 1 and Level 2 students might be artificial in nature – that is, it may be the result of what is commonly known as 'regression to the mean'. The phenomenon of regression to the mean is typically observed when, for example, the lowest-scoring segment of students in respect of a specific test is retested after a while. When these students are retested, their average score may well be higher, not as a result of a real improvement in their ability, but as a result of imperfections in the initial measurement. In the case of this lower segment of students, the ability of a number of students may have been underestimated on the first occasion, which is unlikely to happen on a second occasion. The same phenomenon (but in the opposite direction) may occur when the highest segment of students is retested. However, in our view, it is unlikely that regression to the mean comprises the explanation for the improvement of the Level 1 and Level 2 students, because the test is highly reliable: it has reliability coefficients of well above 0.90. The only way to determine whether regression to the mean possibly provides a feasible explanation in this regard, would be to exclude some of the at-risk students from extra tutoring in academic literacy, which, for obvious, ethical reasons, is not a conceivable option.

As already noted, increased maturity might also explain the improvement of the at-risk students (and again, withholding the currently compulsory extra tutoring from a segment of the at-risk

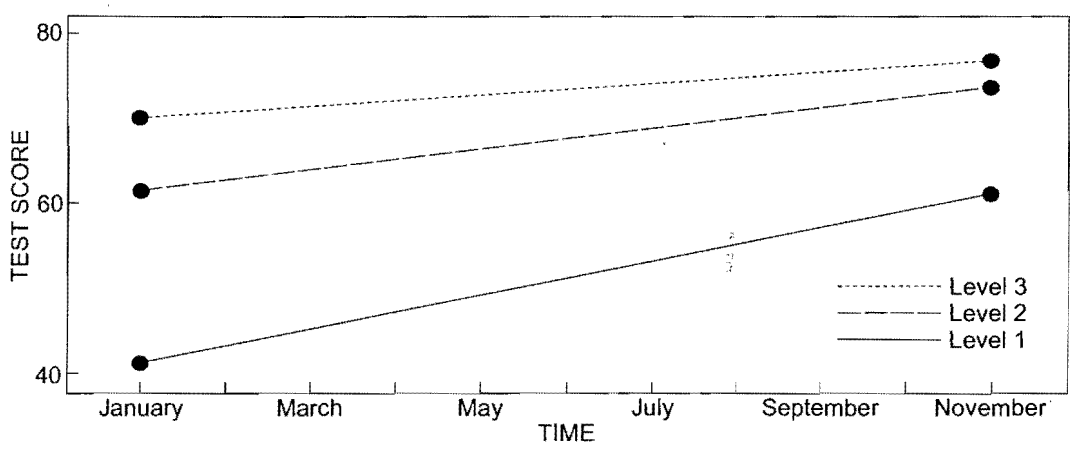

Figure 1: Improvement of academic literacy over time for different initial levels of academic literacy

students could, theoretically, be used to rule out this possibility). However, developmental studies have repeatedly pointed to the existence of the so-called 'Matthew effect' (Merton, 1968).[1] The basic idea is that the 'haves' develop at a faster pace than the 'have-nots'. If this idea is applied to the situation of our students, the logical conclusion would be that, without extra tutoring, exactly the opposite of what has been observed, would occur – that is, the Level 3 students would improve their academic literacy at greater speed than the Level 1 and Level 2 students. Apparently, this is not what has happened – in fact, the opposite was observed – and we therefore contend that the compulsory tutoring in which the group identified in the initial assessment as being at risk (Levels 1–3) is obliged to participate, is the most likely explanation in this regard.

Figure 2 tells us – this time, in keeping with the Matthew effect – that students with mother tongues 1 and 2 (Afrikaans and English) improved their academic literacy between January and November 2005 to a somewhat higher degree than students with mother tongues 3 and 4 (other South African languages and other languages). However, this difference in the improvement rate can only partly be explained by the different dispersion of the level of academic literacy between students whose mother tongue is 1 or 2 and students whose mother tongue is 3 or 4, since the interaction effect of level and mother tongue (see Table 3) is only of borderline significance ($F_{(2, 604)} = 2.88$, $p = 0.057$). It does not seem unreasonable to assume that the languages of instruction at UP are at least partly responsible for this difference in academic literacy improvement.

Sub-tests

What measure of improvement is evident with regard to the six sub-tests that constitute the TALL? It does not seem unreasonable to assume that the improvement did not develop at the same speed in respect of all sub-tests. In order to test this assumption, we set up a 'repeated measures anova' with two within-subjects factors, namely sub-tests, which are nested in Time, and two between-subjects factors, namely level of academic literacy and mother tongue. The descriptive statistics can be found in Appendix 1. In Table 4, we present the outcomes of the analysis of variance.

The assumption that the improvement of academic literacy varies across sub-tests is indeed corroborated by the outcomes listed in Table 4, since the interaction effect Time * Tests is highly significant ($F_{(2,1220)} = 56.68$, $p < 0.001$). Moreover, Table 4 tells us that the improvement of academic literacy over time is not only dependent on the question as to which sub-test is considered. The factors, level of initial academic literacy and mother tongue, play a role as well, since the third-order interaction effect, Time * Tests * Level * mother tongue, is also significant ($F_{(4, 1220)} = 2.74$, $p < 0.027$). In Figure 3, these outcomes are indicated, with the left panel representing the outcomes

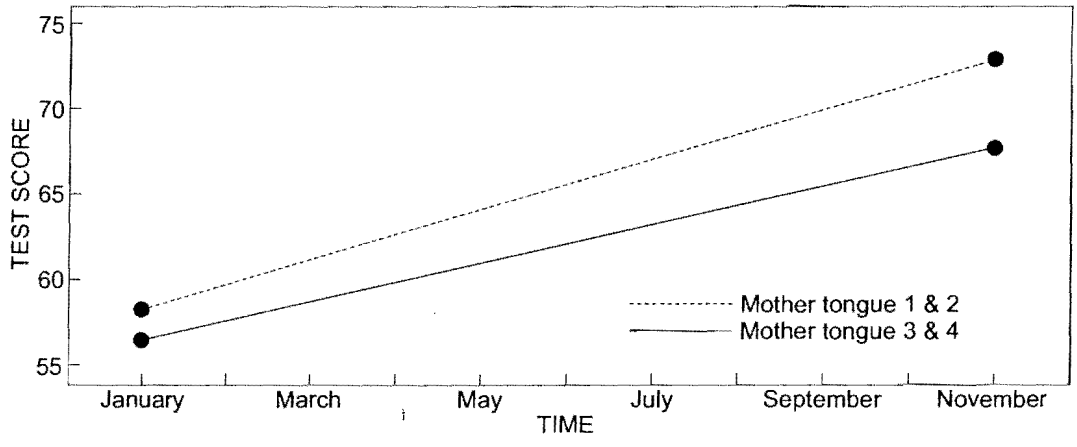

Figure 2: Improvement of academic literacy over time for different mother tongues

Table 4: ANOVA Repeated measures of the January and November TALL test, with level of academic literacy and mother tongue as between-subjects factors and Time and subtest as within-subject factors

	df	Mean Square	*F*	*p*
Time	1	2 801.16	311.16	0.000
Time * level	2	236.69	26.34	0.000
Time * mother tongue	1	67.46	7.51	0.006
Time * level * mother tongue	2	25.51	2.84	0.059
Error (Time)	604	8.99		
Tests	2.1	118 041.25	3 095.80	0.000
Tests * level	4.1	2 448.65	64.22	0.000
Tests * mother tongue	2.1	133.68	3.51	0.029
Tests * level * mother tongue	4.1	89.64	2.35	0.051
Error (Tests)	1 237.7	38.13		
Time * Tests	2.0	1 247.52	56.68	0.000
Time * Tests * level	4.0	236.99	10.77	0.000
Time * Tests * mother tongue	2.0	130.09	5.91	0.003
Time * Tests * level * mother tongue	4.0	60.38	2.74	0.027
(Error) Time * Tests	1 219.8	22.01		

for Afrikaans and English mother tongue students, and the right panel, the outcomes for students who are mother tongue speakers of other languages (note that the *y*-axes are not calibrated on the same scale in each case).

What is immediately apparent when Figure 3 is considered, is that the improvement over time is concentrated in sub-test 2 (Knowledge of academic vocabulary), sub-test 5 (Understanding texts) and sub-test 6 (Text editing). These improvements are greater in the case of students whose low level of academic literacy caused them to be placed in the high-risk category in January. In particular, the improvement displayed by students with an initial Level 1 academic literacy is impressive.

Conclusion

The initial and final scores obtained in the same test in respect of the academic literacy levels of those surveyed here, indicate the extent of the development of their ability to handle academic language. The improvement that is evident is not only highly significant over time, but is also quite dramatic in the case of those who were initially identified as being most at risk (Level 1 students).

However, the downside is that, despite the rapid pace at which the students in the bottom group have improved (see Figure 1), they still have not reached the initially determined lower levels of risk at the end of the year. In this regard, the Level 2 students fare better on average — they do manage to eliminate risk, as measured by the assessment instrument, as of course is the case with regard to Level 3 students.

What does this mean? Firstly, it indicates that what we have been observing over a number of years is indeed the case: there is a 'bottom group' of students of fairly low ability who do not improve sufficiently over time (or at least over the period of the current intervention — one year) to eliminate the risk resulting from too low a level of academic literacy. An investigation should certainly be carried out in order to determine whether it would be economically feasible and worthwhile to proceed with a further intervention that follows on the current one, in order to attempt to rectify the shortfall in ability among these students who are in their second and subsequent years of study. Such a step would certainly not be unthinkable nor undesirable — in fact, we have remarked before about the savings that it would bring about for the higher education system, as well as for individual institutions and for parents, in cases where interventions increase throughput rates. This group should either be given more attention, or the institution that has allowed them access to higher education should tighten its selection criteria. However, in a politically fraught situation such as that in South Africa at present, the latter possibility does not seem to be the best option.

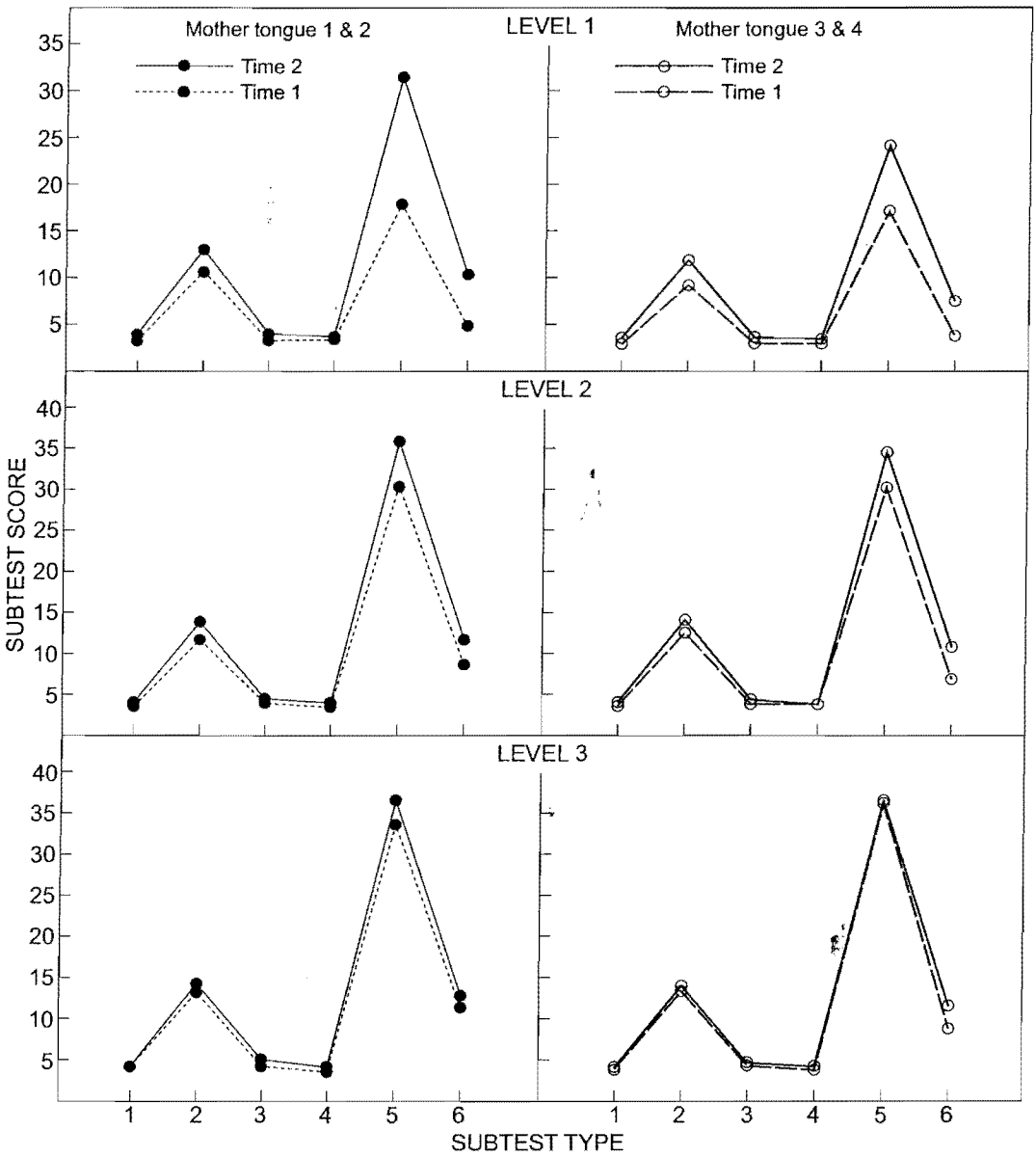

Figure 3: Initial and final sub-test outcomes for mother tongue speakers of Afrikaans and English (left column) and mother tongue speakers of other languages (right column) at different academic literacy levels

Conversely, the analyses on which we are reporting here also show that not only do those at the lowest level of academic literacy need more attention to enable them to develop beyond what can be accomplished in one year, but the bther students initially identified as being at risk (Levels 2 and 3) also benefit significantly from the intervention. Of course, the actual benefit that they derive from achieving a higher level of academic literacy needs to be further tested and demonstrated, preferably in terms of a longitudinal study of their subsequent performance and success at university.

Further study in this respect is necessary, and would be extremely relevant; but the results clearly show the benefits derived from the current intervention.

As we expected, the improvement in academic literacy is uneven across different kinds of tasks, as tested by the various subsections of the test. For example, our observation has been that students do become more familiar with the formats of some sections (of which sub-test 6 is a prime example), which may assist them to eventually achieve higher scores in these sections. However, there is another angle to this finding, which concerns the question that is always at the back of the test designer's mind, namely that of whether the same results and high levels of reliability might not be possible with a shorter test. Some analytical tools, including the analytical instrument we used here, namely Tiaplus (CITO, 2005), even generate statistics indicating what the reliability coefficient might have been had the test been shorter (for example, if 40 rather than 60 items had been included). The fact that the sub-tests in which the students showed the greatest amount of improvement also happen to be the ones whose correlations with the total test score of all students are the highest (see Table 5), means that the use of these 3 sections in a shorter test may be worth considering.

As can be seen from Table 5, the high correlations of sub-test 2, sub-test 5 and sub-test 6 with the total test score – 0.72, 0.95 and 0.79 respectively – are an indication that it may be worthwhile to conduct further tests to determine whether students would be detrimentally affected if they were to write a shorter version of the test, consisting only of these sub-tests. This issue is currently under investigation, and a preliminary analysis and suggestions for future research can be found in Geldenhuys (2007).

The potential advantages of a shorter test are manifold. For example, a shorter test could be used to determine whether a candidate's perception of having been misclassified by the full test — for whatever reason — may be correct. Or it could be used for the purposes of a first, initial screening, with a view to quickly and efficiently identifying the smaller group (in the case of the University of Pretoria, this would refer to approximately the lowest-scoring 30% of those who write the test) whose academic literacy is below the required level, leaving the others who do have the desired level of academic literacy without the burden of having to submit to a longer test; while a full version of the test could be administered to the lowest-scoring group later. That this could serve to further destigmatise the results of the test is obvious. But once again, as we remarked at the beginning, we see here an illustration of how an analysis of the empirical, quantitative properties of a test can enhance its accessibility, transparency and accountability, by providing grounds for arguments on how to ensure that those who submit to a test are treated more fairly. Both the aversion to quantitative analyses that one encounters in some post-modern work in the field of applied linguistics (see Weideman, 2006; Van der Slik & Weideman, 2005) and the prejudice that continues to support this aversion, need to be challenged on this basis. The two concerns — efficiency and effectiveness on the one hand, and accountability and fairness on the other — are complementary, and not contradictory.

Apart from this line of enquiry relating to a shorter test, as suggested by Geldenhuys (2007), the analysis presented here, together with the related discussions that have been referred to above,

Table 5: Sub-test correlations for different sections of TALL 2005 with total test when measured across three different student populations (University of Pretoria, North-West University, and University of Stellenbosch) [n = 5174]

	Sub-test	Total test
Scrambled text	1	0.48
Academic vocabulary	2	0.72
Graphs & visual information	3	0.64
Text types	4	0.41
Understanding texts	5	0.95
Text editing	6	0.79

calls for a number of issues to be addressed in future research. The first, which we have already begun to address in analyses conducted by Van der Slik and Weideman (2007), concerns the equivalence, over several administrations, of one of the tests employed (TAG). This analysis should be extended to TALL; and the equivalence of TALL and TAG themselves should also begin to receive analytical attention.

The second strand of future research required, which Van der Slik (in preparation) is investigating at the moment, and in respect of which he has already provided a preliminary report at a recent SAALA/LSSA conference (2007, Potchefstroom), is related to the possible bias or prejudice that may be inherent in the test. The results of his analysis should be finalised and ready for publication in 2008.

A third issue is the validation of the test. A seminal study on the validation of TAG in the context of its administration in Potchefstroom has been conducted by Van der Walt and Steyn (2007), and needs to be replicated for other administrations. Similarly important are studies of the predictive validity of the test. The analysis by Geldenhuys (2007) has paved the way in this regard. As the 2008 data and results of the tests become available, we shall be able to conduct longitudinal studies, which may be able to tell us whether the anecdotal evidence that we have obtained regarding the ability of the test to predict difficulty and obstacles in academic performance, has any substance or validity.

A fourth concern is the temptation to begin to use this test, which was initially designed as a placement test, as a high stakes test for access purposes. The extremely satisfactory reliability levels (see Table 1, above) make this temptation seem all the more feasible. In our future research, we need to find empirical bases for the current practical condition that we have placed upon the use of the test for high stakes purposes, namely, that it should never be used as the only measure for access, but should always be applied in conjunction with three or four other criteria, which can include aspects as diverse as biographical information, previous academic performance, socio-economic or other factors, aptitude, and so forth.

The final remark that we wish to make in conclusion concerns the variation of the improvement in academic literacy levels across the different first languages (mother tongues). As illustrated by Table 4 and, especially, Table 3, those whose first languages are Afrikaans and English tend to improve much more (in the case of the English version of the test) than those whose first languages are other languages (African and foreign).

The argument concerning this matter can follow one of two directions. First, one may argue that this is a clear sign that the other indigenous South African languages deserve to be affirmed within higher education, if we want the performance of the users of those languages to be on a par with that of their Afrikaans and English counterparts. Indeed, at the University of Limpopo, there is already a dual-medium BA degree in Contemporary English Language and Multilingual Studies (Ramani *et al.*, 2006). Its goal is to promote both knowledge of, and competence in English and Sesotho sa Leboa for academic purposes; and the degree programme is so designed that it uniquely affirms the resources of an indigenous language, and celebrates a commitment to multilingualism.

There is nothing wrong with such an initiative; and critics of this particular programme – who thought that it would have no takers, on the grounds that there is no demand for any language but English – are probably astonished at the growing numbers of students. But there is a second argument, which becomes obscured if the numbers are considered in isolation, namely, that the Afrikaans and English first-language users who performed so well at the University of Pretoria, and who improved so dramatically, were exposed to mother tongue education when the foundations of their academic literacy were being laid. In fact, those speakers of Sesotho sa Leboa who take the new degree course at the University of Limpopo are grappling with another disadvantage: they have been underprepared, by the educational arrangements in the schools that they attended, for performing well in their own first language.

In view thereof, the second argument should be borne in mind as an important motivating factor for the retention of mother tongue education, at least at primary school level and, if possible, the introduction of measures aimed at increasing the usage of the mother tongue as the medium

of instruction at secondary school level as well. Otherwise, generation after generation may be doomed to the kind of underperformance and risk that is all too evident in some of the figures provided above; and the current ability gap in the use of academic language may never be bridged for the benefit of those who are most in need of intervention in this regard.

Notes

[1] Matthew 25:29: For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath.

References

CITO. 2005. *TiaPlus, Classical Test and Item Analysis* © Arnhem: Cito M & R Department.

Cliff AF, Yeld N & Hanslo M. 2006. Assessing the academic literacy skills of entry-level students, using the Placement Test in English for Educational Purposes (PTEEP). Forthcoming in *Assessment in Education.*

Coffin C & Hewings A. 2004. IELTS as preparation for tertiary writing: distinctive interpersonal and textual strategies. In Ravelli LJ & Ellis RA (eds) *Analysing academic writing: contextualized frameworks.* London: Continuum, pp 153–171.

Coxhead A. 2000. A new academic word list. *TESOL Quarterly* **34**(2): 213–238.

Coxhead A & Nation P. 2001. The specialised vocabulary of English for academic purposes. In Flowerdew J & Peacock M (eds) *Research perspectives on English for academic purposes.* Cambridge: Cambridge University Press, pp 252–267.

Davidson F & Cho Y. 2001. Issues in EAP test development: what one institution and its history tell us. In Flowerdew J & Peacock M (eds) *Research perspectives on English for academic purposes.* Cambridge: Cambridge University Press, pp 286–297.

Geldenhuys J. 2007. Test efficiency and utility. Forthcoming in Special Issue of *Ensovoort.*

Jackson PW & Agunwamba CC. 1977. Lower bounds for the reliability of the total score on a test composed of nonhomogeneous items: I. Algebraic lower bounds. *Psychometrica* **42**: pp 567–578.

Merton RK. 1968. The Matthew effect in science. *Science* **159**(3810): 56–63.

Ramani E, Modiba M & Joseph M. 2006. 'Variable competencies in students' bilingual academic writing (in English and Sesotho sa Leboa) in a dual-medium BA degree'. Presentation at the 2006 SAALA conference, Durban, 6 July.

Shohamy E. 2001. *The power of tests: a critical perspective on the uses of language tests.* Harlow: Pearson Education.

Unit for Academic Literacy. 2006. Compulsory academic literacy test. [Available at: http://www.up.ac.za/academic/humanities/eng/eng/unitlangskills/eng/fac.htm Accessed 25 September 2006].

Van der Slik F. In preparation. 'An examination of gender bias in two tests of academic literacy'. Paper presented at the joint LSSA/SAALA/SAALT 2007 conference, Potchefstroom.

Van der Slik, F & Weideman A. 2005. The refinement of a test of academic literacy. *Per Linguam* **21**(1): 23–35.

Van der Slik F & Weideman A. 2007. Testing academic literacy over time: is the academic literacy of first year students deteriorating? Forthcoming in Special Issue of *Ensovoort.*

Van der Walt JL & Steyn HS jnr. 2007. Pragmatic validation of a test of academic literacy at tertiary level. Forthcoming in Special Edition of *Ensovoort.*

Van Dyk T & Weideman A. 2004a. Switching constructs: on the selection of an appropriate blueprint for academic literacy assessment. SAALT *Journal for Language Teaching* **38**(1): 1–13.

Van Dyk T & Weideman A. 2004b. Finding the right measure: from blueprint to specification to item type. SAALT *Journal for Language Teaching* **38**(1): 15–24.

Van Rensburg C & Weideman A. 2002. Language proficiency: current strategies, future remedies. *SAALT Journal for Language Teaching* **36**(1&2): 152–164.

Visser A & Hanslo M. 2005. Approaches to predictive studies: possibilities and challenges. *South African Journal of Higher Education* **19**(6): 1160–1176.

Weideman A. 2003. *Academic literacy: prepare to learn.* Pretoria: Van Schaik.

Weideman A. 2006. Transparency and accountability in applied linguistics. *Southern African*

Linguistics and Applied Language Studies **24**(1): 71–86.
Weideman A & Van der Slik F. 2008. The stability of test design: measuring difference in performance across several administrations of a test of academic literacy. *Acta Academica* **40**(1): 161–182.

Appendix 1: Descriptive statistics of the English version of the six academic literacy sub-tests in January and November 2005 by level of academic literacy in January 2005 and by mother tongue

Sub-test 1		Mother tongue = 1,2			Mother tongue = 3,4			Total		
		Mean	SD	N	Mean	SD	N	Mean	SD	N
January	Level 1	3.27	1.67	22	2.88	1.76	275	2.91	1.75	297
	Level 2	3.55	1.47	84	3.63	1.57	182	3.61	1.54	266
	Level 3	4.12	1.22	17	3.83	1.46	30	3.94	1.37	47
	Total	3.58	1.48	123	3.22	1.72	487	3.29	1.68	610
November	Level 1	3.91	1.51	22	3.57	1.59	275	3.60	1.58	297
	Level 2	4.08	1.29	84	4.06	1.36	84	4.07	1.33	266
	Level 3	4.18	1.38	17	4.03	1.50	30	4.09	1.44	47
	Total	4.07	1.34	123	3.78	1.51	487	3.84	1.48	610

Sub-test 2		Mother tongue = 1,2			Mother tongue = 3,4			Total		
		Mean	SD	N	Mean	SD	N	Mean	SD	N
January	Level 1	10.55	3.04	22	9.19	3.46	275	9.29	3.45	297
	Level 2	11.74	3.22	84	12.53	2.88	182	12.28	3.01	266
	Level 3	13.18	2.83	17	13.27	2.43	30	13.23	2.56	47
	Total	11.72	3.20	123	10.69	3.63	487	10.90	3.57	610
November	Level 1	12.91	2.60	22	11.86	3.33	275	11.93	3.29	297
	Level 2	13.88	2.92	84	14.11	2.97	182	14.04	2.95	266
	Level 3	14.24	3.46	17	13.93	2.65	30	14.04	2.93	47
	Total	13.76	2.95	123	12.82	3.34	487	13.02	3.29	610

Appendix 2: Descriptive statistics of the English version of the six academic literacy sub-tests for January and November 2005 by level of academic literacy in January 2005 and by mother tongue

Sub-test 1		Mother tongue = 1,2			Mother tongue = 3,4			Total		
		Mean	SD	N	Mean	SD	N	Mean	SD	N
January	Level 1	3.27	1.67	22	2.88	1.76	275	2.91	1.75	297
	Level 2	3.55	1.47	84	3.63	1.57	182	3.61	1.54	266
	Level 3	4.12	1.22	17	3.83	1.46	30	3.94	1.37	47
	Total	3.58	1.48	123	3.22	1.72	487	3.29	1.68	610
November	Level 1	3.91	1.51	22	3.57	1.59	275	3.60	1.58	297
	Level 2	4.08	1.29	84	4.06	1.36	84	4.07	1.33	266
	Level 3	4.18	1.38	17	4.03	1.50	30	4.09	1.44	47
	Total	4.07	1.34	123	3.78	1.51	487	3.84	1.48	610

Sub-test 2		Mother tongue = 1,2			Mother tongue = 3,4			Total		
		Mean	SD	N	Mean	SD	N	Mean	SD	N
January	Level 1	10.55	3.04	22	9.19	3.46	275	9.29	3.45	297
	Level 2	11.74	3.22	84	12.53	2.88	182	12.28	3.01	266
	Level 3	13.18	2.83	17	13.27	2.43	30	13.23	2.56	47
	Total	11.72	3.20	123	10.69	3.63	487	10.90	3.57	610
November	Level 1	12.91	2.60	22	11.86	3.33	275	11.93	3.29	297
	Level 2	13.88	2.92	84	14.11	2.97	182	14.04	2.95	266
	Level 3	14.24	3.46	17	13.93	2.65	30	14.04	2.93	47
	Total	13.76	2.95	123	12.82	3.34	487	13.02	3.29	610

Sub-test 3		Mother tongue = 1,2			Mother tongue = 3,4			Total		
		Mean	SD	N	Mean	SD	N	Mean	SD	N
January	Level 1	3.32	1.52	22	2.98	1.51	275	3.01	1.51	297
	Level 2	3.99	1.24	84	3.84	1.30	182	3.88	1.28	266
	Level 3	4.24	1.15	17	4.23	1.04	30	4.23	1.07	47
	Total	3.90	1.30	123	3.38	1.48	487	3.48	1.46	610
November	Level 1	4.00	1.27	22	3.66	1.41	275	3.69	1.40	297
	Level 2	4.54	1.25	84	4.37	1.17	182	4.24	1.20	266
	Level 3	5.06	.93	17	4.60	1.10	30	4.77	1.03	47
	Total	4.51	1.23	123	3.99	1.36	487	4.09	1.35	610

Sub-test 4		Mother tongue = 1,2			Mother tongue = 3,4			Total		
		Mean	SD	N	Mean	SD	N	Mean	SD	N
January	Level 1	3.45	1.18	22	2.99	1.44	275	3.02	1.43	297
	Level 2	3.48	1.27	84	3.81	1.25	182	3.70	1.26	266
	Level 3	3.53	1.01	17	3.73	1.31	30	3.66	1.20	47
	Total	3.48	1.21	123	3.34	1.42	487	3.37	1.38	610
November	Level 1	3.68	1.09	22	3.46	1.41	275	3.48	1.39	297
	Level 2	4.00	1.23	84	3.79	1.36	182	3.88	1.32	266
	Level 3	4.12	1.27	17	4.17	1.29	30	4.15	1.27	47
	Total	3.96	1.21	123	3.63	1.40	487	3.70	1.37	610

Sub-test 5		Mother tongue = 1,2			Mother tongue = 3,4			Total		
		Mean	SD	N	Mean	SD	N	Mean	SD	N
January	Level 1	17.86	8.17	22	17.05	7.33	275	17.11	7.38	297
	Level 2	30.37	5.92	84	30.16	5.50	182	30.22	5.62	266
	Level 3	33.59	3.45	17	36.07	3.64	30	35.17	3.73	47
	Total	28.58	7.96	123	23.13	9.60	487	24.22	9.54	610
November	Level 1	31.45	9.36	22	24.09	8.59	275	24.63	8.44	297
	Level 2	35.80	6.17	84	34.47	7.20	182	34.89	6.91	266
	Level 3	36.53	4.58	17	36.50	5.88	30	36.51	5.40	47
	Total	35.12	6.84	123	28.73	9.55	487	30.02	9.42	610

Sub-test 6		Mother tongue = 1,2			Mother tongue = 3,4			Total		
		Mean	SD	N	Mean	SD	N	Mean	SD	N
January	Level 1	4.81	3.96	22	3.80	2.99	275	3.88	3.07	297
	Level 2	8.65	3.72	84	6.81	3.80	182	7.39	3.86	266
	Level 3	11.29	2.62	17	8.73	3.61	30	9.66	3.48	47
	Total	8.33	4.07	123	5.23	3.74	487	5.86	4.01	610
November	Level 1	10.32	3.50	22	7.47	3.28	275	7.68	3.37	297
	Level 2	11.71	2.81	84	10.76	2.57	182	11.06	2.28	266
	Level 3	12.76	1.20	17	11.47	1.91	30	11.94	1.79	47
	Total	11.61	2.85	123	8.95	3.41	487	9.48	3.47	610