Tobie van Dyk, Alta van Rensburg and **Fiona Marais**

Levelling the playing field: an investigation into the translation of academic literacy tests

A B S T R A C T It is widely accepted that low levels of proficiency in the languages of learning and teaching (in this article, academic language proficiency refers to academic literacy and the terms will be used interchangeably), affect through-put rates negatively. This unsettling trend is confirmed by local and international literature, and can possibly be attributed to the language curriculum in secondary education that does not prepare students adequately for the higher-order language-thinking skills they need for study at university. By this we refer to Bloom's taxonomy, especially the three higher-order skills of analysing, synthesising and evaluating, and the way language is used for these purposes. In order to address this problem, and as part of language-planning initiatives, some faculties at Stellenbosch University introduced the integration of academic literacy courses into the first-year curriculum. These courses are fully credit-bearing and a system of continuous assessment was adopted. Semester tests form part of this assessment process, and led to the investigation done for this paper. Since both Afrikaans- and English-speaking students register for the same academic literacy module it is imperative that outcomes and assessments should be on the same level. However, the aggregate on the Afrikaans semester tests have continuously been lower than on the English test. The aggregate for the Afrikaans tests was, furthermore, on par with the weighted average for all other first-year courses, which was not always the case with the English tests. After an initial investigation, it was concluded that the English-speaking students were not necessarily academically stronger than their Afrikaans counterparts, but it seemed likely that the problem lay with the tests themselves. A first notion was that academic and spoken English are closer than academic and spoken Afrikaans. It was also possible that the level of difficulty of the English test was substantially lower than that of the Afrikaans test. It should, however, be noted that both the Afrikaans and English tests produced excellent reliability coefficients (alpha above 0.88) and most items discriminated adequately. A possible solution to the benchmarking problem was to translate the Afrikaans test into English. The translation framework, adopted for this study, was Nord's functionalist model. This paper will elaborate on the translation procedure, and the variance in students' performance on the translated version compared to previous administrations. Preliminary conclusions on bias in translated tests and the success and feasibility of such procedures are drawn.

OPSOMMING Daar word allerweë aanvaar dat lae vaardigheidsvlakke in die tale van onderrig en leer (in hierdie artikel verwys akademiese taalvaardigheid na akademiese geletterdheid en die twee terme word afwisselend met mekaar gebruik) deurvloeikoerse benadeel. Hierdie onrusbarende tendens word deur plaaslike én internasionale literatuur bevestig, en kan moontlik daaraan toegeskryf word dat die taalkurrikulum in hoërskoolonderrig nie daarin slaag om studente genoegsaam vir die hoërorde-taaldenkvaardighede van universiteitstudie toe te rus nie. Met hoërorde-taaldenkvaardighede word verwys na Bloom se taksonomie, en spesifiek die drie hoërorde-vaardighede van analise, sintese en evaluering, en die wyse waarop taal vir hierdie doel aangewend word. Om hierdie probleem te hanteer, en as deel van taalbeplanningsinisiatiewe, het sommige fakulteite aan die Universiteit Stellenbosch kursusse in Akademiese Geletterdheid by die eerstejaarskurrikulum begin integreer. Hierdie kursusse is ten volle kredietdraend en volg 'n stelsel van voortgesette assessering. Semestertoetse maak deel uit van hierdie assesseringsproses en het tot die ondersoek vir hierdie navorsingstuk aanleiding gegee. Aangesien sowel Afrikaans- as Engelssprekende studente vir dieselfde module in Akademiese Geletterdheid registreer, is dit noodsaaklik dat uitkomste en assessering op gelyke vlak lê. Tog was die totale punt vir die Afrikaanse semestertoets deurgaans laer as dié vir die Engelse toets. Voorts was die totale punt vir die Afrikaanse toets in lyn met die geweegde gemiddelde vir alle ander eerstejaarskursusse, wat weer nie altyd die geval met die Engelse toets was nie. Ná 'n aanvanklike ondersoek is afgelei dat die Engelssprekende studente nie noodwendig akademies sterker as hul Afrikaanse eweknieë is nie, maar dat die probleem waarskynlik by die toetse self lê. 'n Eerste moontlikheid was dat akademiese en gesproke Engels nader aan mekaar is as akademiese en gesproke Afrikaans. Tweedens kon die moeilikheidsgraad van die Engelse toets ook aansienlik laer wees as dié van die Afrikaanse toets. Dit is egter belangrik om daarop te let dat die Afrikaanse én Engelse toetse albei uitstekende betroubaarheidskoëffisiënte (alfa bo 0.88) sowel as merendeels goeie diskrimineringskoëffisiënte opgelewer het. 'n Moontlike oplossing vir die probleem van rigpuntstelling was om die Afrikaanse toets in Engels te vertaal. Die vertaalraamwerk wat vir hierdie studie aanvaar is, was Nord se funksionalistiese model. Hierdie navorsingstuk wei uit oor die vertaalprosedure sowel as die wisseling in studenteprestasie in die vertaalde toets in vergelyking met vorige toetse. Voorlopige gevolgtrekkings word ook gemaak oor sydigheid in vertaalde toetse, en die sukses en uitvoerbaarheid van sodanige prosedures.

Keywords: academic literacy test; functionalist translation approach; back-translation; adaptation of tests

1. Introduction

It is common knowledge that universities around the globe are under considerable pressure to increase student throughput rates. They, therefore, out of necessity, have had to generate ways to advance academic success. This is, however, easier said than done in South Africa, a country where throughput rates fall well below the international norm (Scott, Yeld & Hendry, 2007). There are, of course, many factors contributing to the lack of academic success at South African universities: under-preparedness for university study and difficulties with the transition from school to higher education are, among others, often listed as the main reasons.

Underscoring the reasons listed above, is the fact that students arrive at institutions of higher education with certain literacy practices that are frequently considered inappropriate, or even unacceptable, for the academic context. These practices are often referred to as academic literacy, which, for the purposes of this paper, refers to proficiency in the language(s) of teaching and learning, as well as those reading and writing abilities required for successful study. The authors are aware that this could be considered a simplified definition of the term 'academic literacy': aspects such as numeracy, information literacy, digital literacy, and especially the beliefs, attitudes and values of the individual and/or the group should, in a broader discussion of the topic, form part of the definition (cf. Gee, 1990; 1992; Bourdieu, Passeron & De Saint Martin, 1994; and Street, 1995).

Although we agree with claims that academic literacy (within the context of a simplified definition) is a *sine qua non* for academic success (Van Rensburg & Weideman, 2002:153; and McKenna, 2009:8), it has to be said that a stance like this could easily develop into a so-called crisis narrative as Ivanič, Edwards, Barton, Martin-Jones, Fowler, Hughes, Mannion, Miller, Satchwell and Smith (2009:14) label it. The purpose of this article is therefore, neither to play down the significance nor to overstate the urgency of the matter, but rather to present a case, at Stellenbosch University, where the assessment of academic literacy among first-year students is investigated.

2. Problem statement

Academic literacy courses, as fully credit-bearing, stand-alone modules, are gradually being integrated into the current first-year curriculum at Stellenbosch University. Their aim is to adequately prepare students for the higher-order language-thinking skills required for university study. These courses are presented in both Afrikaans and English. Students are placed in relevant programmes by using the Test of Academic Literacy Levels (TALL) and its Afrikaans counterpart, Toets van Akademiese Geletterdheidsvlakke (TAG) – cf. Van Dyk and Weideman (2004) for a discussion of the construct of the test. It should be noted that each faculty employs a different point of departure and/or procedure for placement, based on individual language plans (the implementation plans/strategies for the University's overarching language policy) in accordance with their specific needs.

Since both Afrikaans- and English-speaking (i.e. language of preference) students in the Faculty of Natural Sciences are registered for the same academic literacy course, it is imperative that outcomes and assessments for these language groups should be on the same level. However, the aggregate on the Afrikaans semester tests (a version of TAG and based on the same construct as TALL) have consistently been lower than that of the English (a version of TALL and based on the same construct as TAG). Furthermore, the aggregate of the Afrikaans students for this module (based on continuous assessment and therefore inclusive of marks for smaller tasks/exercises, as well as major assignments in reading and writing) showed a strong and significant correlation ($r=0.84$; $p<0.0001$) with that of TAG, whereas the correlation of the English speakers' weighted aggregate with TALL indicated a lower, although still significant correlation ($r=0.67$; $p<0.0001$).

With the above as point of departure, followed by a provisional investigation into the academic performance of all first-year students, it was concluded that English-speaking students were not

necessarily academically stronger than their Afrikaans counterparts; on the contrary, students with Afrikaans as their first language, perform in general, better at Stellenbosch University. It goes without saying that there are a number of variables influencing study success, only one of which is the language of learning and teaching (LoLT) and at this university, Afrikaans is, for the most part, used as the LoLT. The most likely conclusion was, therefore, that the tests themselves differed in terms of difficulty, although they had both, on previous occasions and with larger samples, produced excellent reliability coefficients (α=0.89 for the Afrikaans, and α=0.94 for the English).

As a possible solution to the benchmarking problem, it was then decided to investigate both the effectiveness and the consequences of translating an Afrikaans version of TAG into English.

3. Research methodology

In order to address the problem of equivalence, as discussed above, the methodology adopted for the case study, reported on in this article, was both qualitative and quantitative. The translation methodology had to be both carefully thought through and thoroughly described, since the complexities of test translation have often led to a lack of sound procedures and well-considered translation methodologies (Hambleton, 2005:vii).

This article will therefore, firstly elaborate on the translation procedure that was followed, as well as report on the quality assurance mechanisms that were put in place to draw preliminary conclusions on the success and feasibility of such procedures. Secondly, it will discuss the variance in student performance on the translated version, compared to previous administrations, in order to make inferences on the equivalence of the two tests/sets of tests. In this regard, Koch (2009:302) states that

> [e]quivalence is a measurement term dealing with the measurement level at which scores of tests that are available in more than one language … can be regarded as comparable. For test scores to be comparable, it has to be demonstrated that the test is not biased.

We, therefore, had to ensure that both tests measured the same construct, that the degrees of reliability were consistent and that they were both administered under the same conditions.

4. Translation methodology

The importance of a proper description of the translation methodology is stressed by researchers such as Ægisdóttir, Gerstein and Çinarbaş (2009:213); Carroll, Holman, Segura-Bartholomew, Bird and Busby (2001:213); and Koller, Aaronson, Blazeby, Bottomley, Dewolf, *et al*. (2007:1814). We therefore, deemed it necessary to give a detailed description of the translation approach and strategies, followed in this study.

In the literature on the development of tests and instruments used in cross-cultural research, the terms 'translation' and 'adaptation' are used interchangeably. Some scholars use 'adaptation' to refer to all the activities needed to assure the validity of the translated instrument with respect to the target reader (Lamoureux-Hérbert & Morin, 2009:62), while they see 'translation' as one of the steps in the process of adapting tests (Hambleton, 2005:4). Some translation experts such as Nord (2005:28), however, prefer not to make a methodological distinction between

translation (in the narrower sense of the word) and adaptation, but include adaptation as a strategy in the concept of translation.

For the purposes of this paper, the translation of a test will refer to the whole process of creating a conceptually equivalent target text, whereas when an adjustment is made to a part of the translated text, it will be viewed as a particular translation strategy, namely adaptation, within the translation process. Moreover, this paper will report on some of the translation practices, with an in-depth discussion of the merits and weaknesses of back-translation, because, even though not suitable for use in this study, it is a widely used translation strategy in the translation of tests. Back-translation is used successfully in translating, for example, a maths test where literal or word-for-word translation is acceptable. However, the authors argue that, when translating an academic literacy test from Afrikaans into English, where literal translation would result in an incorrect target text, back-translation would not be the optimal translation strategy. A procedure for test translation which seemed to be more appropriate as a theoretical framework for this study, was Nord's (1991a; 1991b; 1997a; 1997b; 2002; and 2005) version of the functionalist approach to translation. Finally, the translation process, used in this particular study, will be discussed.

4.1 Back-translation

This is one of the most commonly used and internationally accepted, translation strategies for the translation of tests, as well as psychological, medical and psychometric questionnaires. Sometimes referred to as reverse translation, back-translation is considered to be a means of validating the accuracy of a translation. This validation process hinges on a process of double translation, thus, after a test is translated from the source language into the target language, the translated version is translated back into the original language by a different translator. The original version of the test is then compared with the back-translated version, which will also be in the source language. If these two versions do not appear to be identical, the translated version is adjusted so that it matches the source text.

This translation strategy was proposed by Brislin (1970) and has subsequently been used in several translation protocols such as the so-called Serial Approach of Herrera, DelCampo and Ames (1993). The International Quality of Life Assessment (IQOLA) project adopted a set of procedures including back-translation to translate the SF-36 Health Survey into several languages (cf. Aaronson, Muller, Cohen, Essink-Bot, Fekkes *et al.*, 1998). Back-translation was also used in the strict translation protocol developed by the EuroQol Group to translate the EQ-5D quality of life measure. The process of translation and the determining of validity can be found in Jelsma, Chivaura, De Weerdt & De Cock (2000) and Jelsma, De Cock, De Weerdt, Mielke & Mhundwa (2002).

An advantage of this translation approach is that researchers have some control over the end result of the translated test, especially when they are not familiar with the target language (Ægisdóttir *et al*., 2007:201). In the case of this study, where a multidisciplinary team was involved, this did not pose a problem since all the team members were proficient in both languages.

According to Hambleton (2005:12), the main weakness of the back-translation approach is the high level of inference that translators have to make about the equivalence between the

original and the back-translated version. Furthermore, several studies deem back-translation to be insufficient for validating a translated test (cf. McGorry, 2000; Fourie & Feinauer, 2005; and Lamoureux-Hébert & Morin, 2009). Hambleton (2005:13) supports this by stating that "[e]vidence of test equivalence provided by a back-translation design is only one of many types of evidence that should be compiled in a test adaptation study."

Another of the weaknesses of back-translation is that it usually leads to literal translation at the cost of vernacular language used in the translated version. A literal or word-for-word translation does not guarantee that the correct meaning is transferred from the source text to the translated version. An example from the academic literacy test, translated for this study, is taken from the section dealing with different text types/genre/register where the sentence reads "Is jy reg om te waai?" for which a word-for-word translation into English would be "Are you ready to wave?" Such a translation would certainly have resulted in confusion, since the meaning in Afrikaans is "Are you ready to leave?" Back-translation of the literal translation ("Are you ready to wave?") would likely have resulted in a return to the original sentence, namely "Is jy reg om te waai?" Therefore, with the back-translation procedure, non-vernacular language use and non-transference of the meaning of the original could easily go undetected.

In the case of translated tests, it is particularly important that the target text should strive to be the conceptual equivalent of the source text, using clear and correct language to facilitate successful communication with the target reader. In order to achieve this, the translator needs information such as the purpose of the target text as well as the profile of the target readers, for example their age and literacy levels. A translation theoretical framework which takes all these factors into account, is Nord's version of the functionalist translation approach (Nord, 1991a; 1991b; 1997a; 1997b; 2002; and 2005). The success of the functional translation approach has been proven in studies on the translation of various text types, for example, medical questionnaires (Fourie & Feinauer, 2005), news texts for radio (Van Rooyen & Naudé, 2009), educational and public health material developed for the general public (Feinauer, 2003; Labuschagne & Naudé, 2003; and Colina, 2008 & 2009), and medical texts for professional and academic purposes (Labuschagne & Naudé, 2003). We, therefore, decided to investigate whether this approach would be successful in the translation of tests as well.

4.2 Nord's functionalist translation approach

The key concept of this theory, which originated from Vermeer's Skopos Theory (cf. Vermeer, 1989; 1998), is that the purpose (*skopos*) of the translation plays a bigger role in determining the translation method than the nature of the source text. This means that where a literal translation would not lead to a successful transfer of meaning from source text to target text, the translator can apply a translation strategy, such as adaptation, in order to fulfil the purpose of the target text.

In her own version of the functionalist approach, Nord added the concept of loyalty, which highlights the "responsibility translators have toward their partners in translational interaction" (Nord, 1997b:125). In this study it means, firstly, that the translator must be loyal to the test developer and the source text (the original Afrikaans test being translated), by transferring the correct meaning from the source text to the target text (the translated version). Secondly,

it means that the translator must be loyal to the students taking the translated test (the target readers), by creating an authentic (English) text that does not read like a translation. The translator shows his/her loyalty to the test developer as well as the students by applying appropriate translation strategies, for example, adapting parts where a literal translation would result in an inaccurate translation. The translator and the test developer are partners in the sense that they work together to create a successful translation. The test developer, for instance, provides the translator with a 'translation brief' in which information such as the purpose of the translation and the profile of the target reader are specified, while the translator uses his or her expertise in translation to create a functionally equivalent test for the target reader.

The success of the functionalist approach described in studies on other text types such as medical questionnaires (Fourie & Feinauer, 2005) as well as the additional concept of loyalty, explained above, informed our decision to use Nord's version of the functionalist translation approach as the theoretical framework within which to translate the academic literacy test.

5. An academic literacy test: the translation process

An example of Nord's concept of loyalty, where the translator and test developer work together, is that a multidisciplinary team is responsible for the translation process. This quality assurance procedure was also followed in other studies described in for example Carroll *et al.* (2001); and Koch (2009). The multidisciplinary team for the current study, consisted of test, as well as translation experts, all of whom were bilingual, namely

- a professional translator with twenty years' experience;
- a professional translator with postgraduate qualifications in Linguistics and in the process of completing a postgraduate qualification in Translation Studies with five years' experience;
- a professional translator with a postgraduate qualification in Translation Studies with ten years' experience;
- an academic literacy lecturer with experience in academic development programmes and language testing as well as a postgraduate qualification in Linguistics;
- an academic literacy and testing expert with postgraduate qualifications in Education and Linguistics with thirteen years' experience; and
- an academic literacy lecturer with experience in academic development programmes with a postgraduate qualification in English and in the process of completing a postgraduate qualification in Linguistics.

The translation brief, given to the translator by the test developer, specified the purpose of the translated academic literacy semester test to be a means of measuring the academic literacy levels of the target readers, namely English-speaking first-year students in the Faculty of Natural Sciences. The translation strategy was to provide a faithful translation of the source text, while adapting parts and/or items where necessary. The aim was to create an English translation that would be conceptually equivalent to, and with a similar level of difficulty as, the original Afrikaans test.

The translation process began with an initial translation by a competent second language English speaker with extensive translation experience, followed by a quality control check

by another translator. The test was then reviewed and edited by a mother-tongue English speaker with experience in designing and developing language tests. The changes made to the translation were mostly for reasons of internal validity, since issues such as lack of clarity can lead to measurement error, which clearly has a direct impact on the test scores. In order to ensure clarity, the test instructions received particular attention: on occasion literal translation or incorrect terminology could be a cause of confusion, thereby influencing the performance of the test-taker. Bearing in mind that this was a test of academic literacy, choice of vocabulary was also an important factor and several changes were made to make the text more formal and academic in terms of style and register. Correct grammar, including appropriate syntax and use of tenses, as well as accurate word choice, an important criterion for an academic text, was another consideration and motivated certain alterations to the translation. From a testing perspective, the distractors in a multiple choice test should, ideally, be more or less the same length, whether they are words or sentences, and this informed the rationale for such changes, where necessary.

This second version was then discussed by the multidisciplinary team at a workshop where the English translation was compared to the original Afrikaans test. This was especially beneficial, since it afforded both the translation and the test experts an opportunity to share views and prepare a final translation. There was general consensus that all potential ambiguity be removed from questions and that effective communication be established with the test-taker. Furthermore, as has been mentioned earlier, it is of the utmost importance that the two tests be as equal as possible in terms of difficulty.

6. Test equivalence

Test equivalence refers, within the context of this article, to a situation where a particular test is available for administration to two different populations, who may differ in both language and culture, to make sure that test scores can be effectively compared. Simply put, it refers to the similarity or the relationship between two or more forms of the same test. The object of this, in essence, is to enhance fairness among individuals and groups (Cook & Schmitt-Cascallar, 2005:139). Koch (2009:307) claims that "[t]he main purpose for analysing the equivalence of different language versions of tests are … to investigate the question as to whether one can with a reasonable amount of confidence interpret and use the test scores of the two versions in the same manner."

As was implied in the previous paragraph, the reasons for translating tests are normally quite clear. However, Hambleton (2005:4) claims that "methods and guidelines for preparing test adaptations and establishing the equivalence of scores are not well known." A possible reason for this, also postulated by Hambleton (2005:4), is that "some cross-cultural researchers have even suggested that a high percentage of the research in their field is flawed to the point of being invalid because of poorly adapted tests". Another issue is the notion that familiarity with both translation AND equating procedures are not the case here, since expertise in testing and/or translation does not always reside with the same person. This article is, therefore, also an attempt to bring together these fields of expertise and, to a certain extent, emphasise the importance of a multidisciplinary approach by employing a team of specialists, as was the case in this study. The translation strategies and processes followed in this study have already been

discussed in detail in the previous section, so the focus will now turn to a theoretical discussion on the establishment of equivalence, or test comparability. This will then be followed by a discussion of the statistical data obtained for the purposes of equating the two tests.

Davies, Brown, Elder, Hill, Lumley and McNamara (1999:198) emphasise that equivalent forms of tests should be based on the same construct and set of test specifications, as was the case in the investigation carried out for this article. Scores would, therefore, be expected to be equivalent and the means and variance should also theoretically be equal, or as Koch (2009:306) puts it, "[e]quivalence deals with the measurement or scale level at which scores can be compared across groups." However, Davies et al. (1999) also maintain that since a test is compiled for a specific population, for a specific purpose, it may assess language skills and draw inferences which are different from another test with a similar construct.

In this particular study, different approaches could have been followed to equate test results, but given the context and ethical considerations, a specific approach had to be adopted where the translated test (that is the adapted test) was administered to a part of the sample – first-year BSc students with English as language of preference. We had to ensure that the construct of the two tests being measured was as close as possible (note that the construct may differ slightly as a result of variance in the structure of the languages). Other considerations include the fact that test reliability should be completely even (the forms could thus be interchangeable), and that the conditions under which both tests were written, were similar. Failure to meet these considerations could cause problems of construct, method, and item bias (Koch, 2009:306) to arise. Neither construct nor method bias were applicable to our case study, but the third consideration could have had an influence on differences between the two (language) groups.

Group differences were seemingly not particularly significant in this case, since construct and method bias were minimised as far as possible. However, they still had to be explored with regard to reliability, mean score differences and Differential Item Functioning (DIF) including differences in terms of item properties. Koch (2009:308) states that

> [g]roup differences are not regarded as intrinsically problematic as they may be an indication of real differences on the construct of interest. There is, however, a growing recognition in the international testing community that large group differences in terms of test scores need to be explored extensively for bias before these score differences can be accepted as a true reflection of differences in ability on the construct of interest.

6.1 Differences in terms of reliability

According to Van der Walt and Steyn (2007:143) "[a] completely reliable test implies that test scores are free from errors and can be depended on for making [certain] decisions." In low-stakes tests, such as those used for this study – namely, proficiency tests, lower alpha values are acceptable while for high stakes tests, like access tests, the generally accepted norm is 0.8 (Weir, 2005:29). It should be noted that at Stellenbosch University, TAG and TALL are used for placement purposes as well as part of the access test battery and therefore can be considered as medium-stakes tests at this institution. One should, however, keep in mind that the quality of a test (in this case most specifically a translated test) is not reflected by its reliability coefficient (Van der Walt & Steyn, 2007:143). The latter is merely an indication of its internal consistency,

and for this study, it was necessary to ensure that there were no significant differences in terms of reliability between the two tests. The respective measures of internal consistency for (i) the test as a whole, and (ii) the different sections of the tests were consequently ascertained and the results of the September term tests (the first instance where a translated version of a test was used) are displayed in Table 1.

Table 1 Coefficient alpha for the September term tests

	Test	1 Scrambled text (N = 5)	2 Graphic and visual (N = 5)	3 Text types (N = 5)	4 Compre-hension (N = 25)	5 Academic vocabulary (N = 20)	6 Text relations (N = 12)
Afrikaans	0.70	0.82	0.21	0.69	0.46	0.47	0.69
English	0.68	0.55	0.18	0.70	0.48	0.37	0.71

When interpreting the above, it is clear that the overall measures of reliability for the two tests were fairly close and at an acceptable level. This is especially true if one considers that the results for these tests are only used as part of the continuous assessment of a course and do not contribute more than 6.25% of the final mark for the specific course. The lower values recorded in some of the subsections were to be expected perhaps as a result of a difference in the number of questions for those particular sections. Van der Walt and Steyn (2007:144), agree that alpha coefficients are indeed influenced by the number of items they are based on. What is of concern though, is the larger than expected difference between the two tests for sections 1 (Scrambled text) and 5 (Academic vocabulary). This needs further investigation but is at present beyond the scope of this article. It is, however, worthwhile to take note of the work of Hunter and Schmidt (2000:151), and Kline (2004:559), all of whom argue that it is better, when investigating possible bias, to interpret the measures of the test as a whole rather than those of the various subsections. Since inquiries into measurement bias usually take place at test level, this seems to correspond with, and substantiate this argument. Furthermore, if one considers the outcome of the calculation $(1\text{-alpha}_1)/(1\text{-alpha}_2)$, it is clear that there is no concern for the tests differing in reliability.

6.2 Differences in terms of mean scores

Differences in mean scores were determined by performing T-tests to control whether students who completed the English test (the translated one) performed significantly differently from those who had completed the Afrikaans test. The results of the May, as well as the September semester tests were used. It is important to note that a translated test was only used in September and not for the May semester test. In May, the Afrikaans and English tests were completely different in terms of content, but were both based on the same construct. Figures 1 and 2, below, are graphical representations of the mean score differences for May and September, respectively.

From the following, it is clear that the discrepancy in mean score differences was reduced in the September tests. There may, however, still be cause for concern, as the difference between the two tests remained statistically significant ($p<0.95$). Tables 2 and 3, below, are the descriptive statistics for the two tests.

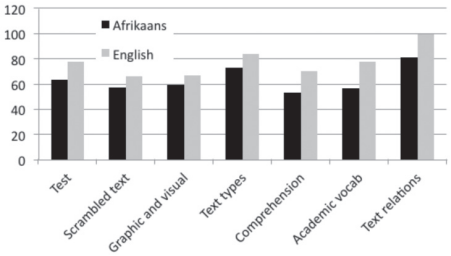

Figure 1 Mean score differences for May 2010

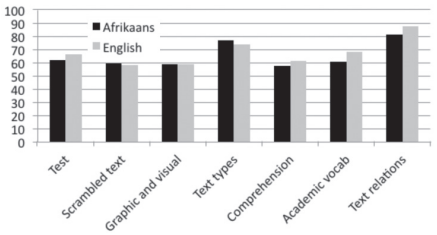

Figure 2 Mean score differences for September 2010

Table 2 Descriptive statistics for performance on the May and September Afrikaans tests (N=235)

		Average	Std Dev.	Difference	t	p
Scrambled text	May	43.49	31.65			
Scrambled text	September	60.43	34.37	-16.94	-5.693	0.00000
Graphic and visual	May	69.69	11.12			
Graphic and visual	September	59.57	22.90	10.11	6.899	0.00000
Text types	May	85.19	24.64			
Text types	September	77.45	25.09	7.74	3.831	0.00016
Comprehension	May	56.88	14.49			
Comprehension	September	61.62	18.51	-4.74	-3.607	0.00038
Academic vocabulary	May	60.46	11.75			
Academic vocabulary	September	58.20	12.28	2.27	2.618	0.00943
Text relations	May	89.16	10.31			
Text relations	September	81.71	14.45	7.46	7.976	0.00000

Table 3 Descriptive statistics for performance on the May and September (translated) English tests (N=235)

		Average	Std Dev.	Difference	t	p
Scrambled text	May	62.72	38.47			
Scrambled text	September	58.64	23.35	4.09	1.3895	**0.16599**
Graphic and visual	May	77.20	15.60			
Graphic and visual	September	58.47	21.27	18.73	12.2820	0.00000
Text types	May	90.30	19.69			
Text types	September	74.64	28.35	15.66	8.3811	0.00000
Comprehension	May	71.88	11.14			
Comprehension	September	61.78	12.43	10.09	13.4725	0.00000

		Average	Std Dev.	Difference	t	p
Academic vocabulary	May	62.09	16.63			
Academic vocabulary	September	67.49	16.80	-5.40	-4.3768	0.00002
Text relations	May	91.17	11.48			
Text relations	September	87.24	14.95	3.93	4.0197	0.00008

These results indicate statistically significant differences in all cases, except for the Scrambled text section in the translated version (the September English test). The null hypothesis could thus, not be rejected. However, to return to the previous argument on whether a test should be analysed as a whole rather than as subscales, it can be concluded that the null hypothesis was indeed rejected and that there were significant differences between the May and the September tests.

6.3 Differential Item Functioning (DIF), including differences in terms of item properties

In short, Differential Item Functioning refers to bias, and, according to Koch (2009:306), comprises "nuisance factors at an item level [contributing to a difference in functioning on the item level]." These factors could cause the performance of a test-taker to depend on abilities other than the ones being tested. After comparing the Afrikaans test and its translated counterpart for the September test, it was concluded that only ten items showed DIF in terms of item difficulty and discrimination index: two were from the Scrambled text section (1 English, 1 Afrikaans), one from the Graphic and visual section (English only), two from Text types (1 English, 1 Afrikaans), three from the Comprehension (English only), and two from the Text relations section (1 English, 1 Afrikaans). The TiaPlus (CITO 2005) statistical package was used to perform the Mantel-Haenszel statistic for the above-mentioned tests and the results are shown below.

Table 4 Descriptive statistics for items showing DIF on the September tests

Item	DIF stat	z(stand)	More difficult for	Section
2	0.2129	-2.7703	Afrikaans	Scrambled text
5	3.6303	2.6242	English	Scrambled text
6	3.452	2.7062	English	Graphic and visual
13	4.1476	2.619	English	Text types
15	0.0996	-2.9552	Afrikaans	Text types
28	3.2964	2.6953	English	Comprehension
29	5.8229	3.6271	English	Comprehension
37	3.952	2.9054	English	Comprehension
56	7.6295	2.6788	English	Text relations
61	0.1658	-3.3855	Afrikaans	Text relations

When interpreting the above, it should be noted that when the DIF statistic is < 1 then the s item proved more difficult for the first subgroup. If the DIF statistic is approximately 1 then the item was equally difficult for both subgroups. If the DIF statistic is > 1 then the item was more difficult for the second subgroup. Note that DIF is only shown at a significance level of alpha = 1%; for alpha = 5% there will be more items showing DIF. Differences between the subgroups are significant when the absolute value of z (stand) ≥ 2.58. Interestingly enough, although there were more questions that seemed to disadvantage the students who wrote the English test, they still performed better overall, than the Afrikaans students. It is possible to speculate on each of the ten statistics shown in Table 4, but for this article, we will be limiting our discussion to Question 6, from the Graphic and visual section of the test.

In Afrikaans, the question read as follows:

> *In watter tydperk is daar 'n ooreenstemming tussen die tendense van die sosiale en geesteswetenskappe en gesondheidswetenskappe?*
>
> *A. 2004-2006*
>
> *B. 2002-2004*
>
> *C. 2004-2007*
>
> *D. 2002-2005*

In English, it read as follows:

> *Which period shows a corresponding trend between the Social Sciences and Humanities and the Health Sciences?*
>
> *A. 2004 and 2006*
>
> *B. 2002 and 2004*
>
> *C. 2004 and 2007*
>
> *D. 2002 and 2005*

The statistics showed that 48.6% of the Afrikaans students, but only 24.5% of the English answered the question correctly – the DIF statistic was 3.452 for this particular question, which indicated that there was a definite problem. When we compared the English translation with the Afrikaans test, we found a translation error: the hyphen, present in all the distractors, signifying the word 'to', had been translated as 'and' which understandably confused the English test-takers, because the distractor referred to a period and not to specific years. Note: only one translation error was made, and that was at question six. The other problematic items can be attributed to general test challenges, for example distractors (the order they appear in and their length), phrasing of questions, and use of terminology – these, however, are beyond the scope of this article and will not be discussed here. Table 5 is an indication of the differences in item properties (mean item difficulty and discrimination values) for the items that showed DIF.

Table 5 Differences in item properties for items showing DIF on the September tests

Item	Afrikaans % correct	English % correct	Afrikaans Discrimination index	English Discrimination index
2	61.8%	89.8%	29.2%	16.1%
5	42.6%	21.9%	38.4%	27.4%
6	48.6%	24.5%	36.5%	9.5%
13	85.5%	64.6%	22.3%	18.8%
15	57.0%	74.5%	35.1%	23.0%
28	68.7%	43.4%	12.8%	16.2%
29	72.3%	38.7%	34.5%	36.2%
37	52.6%	28.5%	49.5%	26.5%
56	94.0%	75.9%	7.4%	35.7%
61	52.6%	88.0%	40.9%	12.0%

7. Conclusion

The reduced mean result of the English September semester test in comparison to previous English tests seems to indicate that the problem of equivalence between the English and the Afrikaans tests is being addressed. It, furthermore, attests to the success of the translation process, except for the one translation error, identified by the statistics and discussed above. Had the test been translated by two independent translators, the error would certainly have been identified earlier on, but would have made the process much more costly. This shows how important it is for a translation team to be particularly vigilant about checking the translation against the original. What has emerged, is that more research needs to be done on the differences that exist in academic literacy assessment in the various languages and the impact on achievement. Further study, involving an alternative to the functionalist translation methodology, such as back-translation, in the translation of academic literacy tests, could also be examined in the future. However, instead of striving for test equivalence by means of translation, and possibly lowering reliability levels, perhaps one should focus on the language and cognitive abilities that one wishes to test, which may differ from language to language. This begs the question of whether the time and money, spent on translation, could be better spent on improved test design.

BIBLIOGRAPHY

Aaronson, N., Muller, M., Cohen, P., Essink-Bot, M., Fekkes, M., Sanderman, R., Sprangers, M., Te Velde, A. & Verrips, E. 1998. Translation, validation, and norming of the Dutch language version of the SF-36 health survey in community and chronic disease populations. *Journal of Clinical Epidemiology*, 51(11):1055-1068.

Ægisdóttir, S., Gerstein, L. & Çinarbaş, D. 2009. Methodological issues in cross-cultural counseling research: Equivalence, bias and translations. *The Counseling Psychologist*, 36(2):188-219.

Bourdieu, P., Passeron, J. & De Saint Martin, M. 1994. *Academic discourse*. Cambridge: Polity Press.

Brislin, R.W. 1970. Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1:185-216.

Carroll, J., Holman, T., Segura-Bartholomew, G., Bird, M. & Busby, D. 2001. Translation and validation of the Spanish version of the RELATE questionnaire using a modified Serial Approach for cross-cultural translation. *Family Process*, 40(2):211-231.

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

Colina, S. 2008. Translation quality evaluation. *The Translator*, 14(1):97-134.

Colina, S. 2009. Further evidence for a functionalist approach to translation quality evaluation. *Target*, 21(2):235-264.

Cook, L.L. & Schmitt-Cascallar, A.P. 2005. Establishing score comparability for tests given in different languages. Pp. 139-169 in Hambleton, R.K., Merenda, P.F. & Spielberger, C.D. *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T. & McNamara, T. 1999. *Dictionary of language testing.* (Series editor: Milanovic, M.). Cambridge: Cambridge University Press.

Feinauer, I. 2003. TB or not TB? The communicative success of translated medical texts in South Africa. In W. Botha (ed). 'n Man wat beur: *'n Huldigingsbundel vir Dirk van Schalkwyk.* Stellenbosch: Buro van die WAT. 213-223.

Fourie, J. & Feinauer, I. 2005. The quality of translated medical research questionnaires. *Southern African Linguistics and Applied Language Studies*, 23(4):349-367.

Gee, J. 1990. *Social linguistics and literacies: ideology in discours*e. London: Falmer.

Gee, J. 1992. *The social mind: Language, ideology and social practices*. New York: Bergin and Garvey.

Hambleton, R.K. 2005. Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. Pp. 3-38 in Hambleton, R.K., Merenda, P.F. & Spielberger, C.D. *Adapting educational and psychological tests for cross-cultural assessment.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Herrera, R., DelCampo, R. & Ames, M. 1993. A serial approach for translating family science instrumentation. *Family Relations*, 42:357-360.

Hunter, J.E. & Schmidt, F.L. 2000. Racial and gender bias in ability and achievement tests. *Psychology, Public Policy, and Law,* 6(1):151-158.

Ivanič, R., Edwards, R., Barton, D., Martin-Jones, M., Fowler, Z., Hughes, B., Mannion, G., Miller, K., Satchwell, C. & Smith, J. 2009. *Improving learning in college: Rethinking literacies across the curriculum.* Oxon: Routledge.

Jelsma, J., Chivaura, V., De Weerdt, W. & De Cock, P. 2000. A bridge between cultures: a report on the process of translating the EQ-5D instrument into Shona. *SA Journal of Physiotherapy*, 56(4):3-9.

Jelsma, J., De Cock, P., De Weerdt, W., Mielke, J. & Mhundwa, K. 2002. The validity of the Shona version of the EQ-5D quality of life measure. *SA Journal of Physiotherapy*, 58(3):8-12.

Kline, T.J.B. 2004. Gender and language differences on the Test of Workplace Essential Skills: Using overall mean scores and item-level Differential Item Functioning analyses. *Educational and Psychological Measurement*, 64(3):549-559.

Koch, E. 2009. The case for bilingual language tests: a study of test adaptation and analysis. *South African linguistics and Applied Language Studies*, 27(3):301-316.

Koller, M., Aaronson, N., Blazeby, J., Bottomley, A., Dewolf, L., Fayers, P., Johnson, C., Ramage, J., Scott, N. & West, K. 2007. Translation procedures for standardised quality of life questionnaires: The European Organisation for Research and Treatment of Cancer (EORTC) approach. *European Journal of Cancer*, 43:1810-1820.

Labuschagne, C. & Naudé, J. 2003. 'n Funksionalistiese benadering tot mediese vertaling in Suid-Afrika. *Acta Academica Supplementum*, 2:129-167.

Lamoureux-Hébert, M. & Morin, D. 2009. Translation and cultural adaptation of the supports intensity scale in French. *American Journal on Intellectual and Developmental Disabilities*, 114(1):61-66.

McGorry, S. 2000. Measurement in a cross-cultural environment: Survey translation issues. *Qualitative Market Research: An International Journal*, 3(2):74-81.

McKenna, S. 2009. Cracking the code of academic literacy: An ideological task. Pp. 8-15 in C. Hutchings & J. Garraway (eds.). *Beyond the university gates: Provision of Extended Curriculum Programmes in South Africa.* Proceedings of the January 2009 Rhodes University Foundation Seminar.

Nord, C. 1991a. Scopos, loyalty, and translational conventions. *Target*, 3(1):91-109.

Nord, C. 1991b. *Text analysis in translation*. Amsterdam: Rodopi.

Nord, C. 1997a. A functional typology of translations. In Trosborg, Anna (ed). *Text typology and translation*. Amsterdam: John Benjamins. 43-66.

Nord, C. 1997b. *Translating as a purposeful activity: Functionalist approaches explained.* Manchester: St Jerome.

Nord, C. 2002. Manipulation and loyalty in functional translation. *Current Writing*, 4(2):32-44.

Nord, C. 2005. *Text analysis in translation: theory, methodology, and didactic application of a model for translation-oriented text analysis.* Second edition. Amsterdam: Rodopi.

Scott, I., Yeld, N. & Hendry, J. 2007. A case for improving teaching and learning in South African higher education. *Higher Education Monitor:* 6. Pretoria: Council for Higher Education.

Street, B. 1995. *Social literacies: Critical approaches to literacy in development, ethnography and education.* London: Longman.

Van der Walt, J.L. & Steyn, H.S. (Jnr). 2007. Pragmatic validation of a test of academic literacy at tertiary level. *Ensovoort*, 11(2):138-153.

Van Dyk, T.J. & Weideman, A.J. 2004. Switching constructs: On the selection of an appropriate blueprint for academic literacy assessment. *Journal for Language Teaching*, 38(1): 1-13.

Van Rensburg, C. & Weideman, A.J. 2002. Language proficiency: Current strategies, future remedies. *Journal for Language Teaching*, 36(1&2):152-163.

Van Rooyen, M. & Naudé, J. 2009. A model for the translation of news agency tests (Sapa) for radio (OFM) news. *Communicatio*, 35(2):251-275.

Vermeer, H. 1989. Skopos and commission in translational action. Translated by Andrew Chesterman. In Venuti, Lawrence (red). 2000. *The Translation Studies Reader*. London: Routledge. 221-232.

Vermeer, H. 1998. Starting to unask what translatology is about. *Target*, 10(1):41-68.

Weir, C.J. 2005. *Language testing and validation*. Houndmills, Basingstoke: Palgrave Macmillan.

ABOUT THE AUTHORS

Tobie van Dyk, of Stellenbosch University's Language Centre, has been working intensively on different kinds of language tests designed for different purposes since 2004. His expertise, however, lies in the field of academic literacy testing and development. He currently heads up the Unit for Afrikaans and English in the Language Centre which is, among others responsible for all language testing, as well as academic literacy development and the acquisition of Afrikaans and English at this university. In addition, he administers the collaborative effort of four local universities, known as the Inter-Istitutional Centre for Language Development and Assessment (ICELDA).

Email: tjvd@sun.ac.za

Alta van Rensburg is head of the Language Service at the Stellenbosch University Language Centre. As part of her doctoral studies, she explores a process approach to quality in translation, focussing on study material as a genre in the context of a higher education institution.

Email: avrens@sun.ac.za

Fiona Marais is with the Unit for Afrikaans and English at Stellenbosch University's Language Centre. She is involved with course design and lecturing in the field of academic literacy. She is particularly interested in language assessment and testing. She developed a computerised academic listening test as part of her Master's study.

Email: fcm@sun.ac.za