

**Ilse Fouché**  
University of Pretoria

# Impact measurement: quantitatively determining the improvement in students' academic literacy levels at a South African university

## Abstract

Academic literacy interventions are becoming increasingly important in a country where the secondary education system no longer adequately prepares students for the literacy demands of higher education. This article investigates whether there was an improvement in students' academic literacy levels between the onset and completion of an academic literacy module at a South African university. This is done by using a combination of instruments selected from a proposed evaluation design for academic literacy interventions, suggested by Fouché, Van Dyk and Butler (2016). A pre-test / post-test design is used, where, firstly, students' results in a validated and reliable generic academic literacy test are considered. Secondly, students' writing abilities are assessed by means of two instruments: a rubric and quantitative measures. Finally, students' academic literacy

marks are correlated to other variables, and interpreted within the context of the study, to give additional insight into the impact of the academic literacy course. Findings indicate that students showed an improvement across a wide array of academic literacy abilities, in particular their ability to use source material in their writing assignments, and their usage of a wider range of academic vocabulary. However, there were also areas where students did not display any improvement. Based on the experience of implementing various evaluation instruments, several recommendations are made on how future researchers could avoid pitfalls that were encountered in this study.

**Keywords:** Academic literacy, English for Academic Purposes, Assessment, Impact

## 1. Introduction

Yeld (2010: 26) states that in recent years, there has been an increasing focus on graduation rates, as well as the effectiveness and quality of higher education. To assist in improving graduation rates, and to provide evidence of their effectiveness, academic literacy interventions firstly have the responsibility of indicating how they contribute to preparing students for the literacy demands of higher education. Secondly, academic literacy interventions have the responsibility to show what measurable impact they have on students' academic literacy levels. It is also important to determine academic literacy competencies on which academic literacy interventions do not have a sufficient impact so that recurriculation can occur, thus ensuring that academic literacy interventions are optimally effective (Newcomer & Triplett, 2010: 6). In fact, the evaluation process should ideally be seen as a continuing needs assessment that constantly works towards improving the academic literacy intervention in question (Bachman & Palmer, 2010: 25; Brown, 2001: 15).

The current study defines impact (or effect, which is used synonymously in this article) as “i) the observable improvement in academic literacy abilities between the onset and the completion of an academic literacy intervention and ii) the extent to which these abilities are necessary and applied in students' content subjects” (Fouché, 2015: 21). This article only considers the first aspect, namely measuring an observable improvement in students' academic literacy levels. This aspect, however, is investigated by using several methods.

This article forms part of a larger study in which an evaluation design is proposed which can be used to assess the impact of a variety of academic literacy interventions. Two previous articles have, firstly, considered national and international studies that have aimed to assess the impact of academic literacy interventions in the past (Fouché, 2015), and secondly, proposed a flexible, comprehensive design that could be used to validly and reliably assess the effectiveness of a variety of academic literacy interventions, including discipline-specific interventions, generic interventions, writing centre interventions as well as reading programmes (Fouché, Van Dyk & Butler, 2016). This article uses a variety of instruments proposed in Fouché et al. (2016) so as to quantitatively assess the impact of a specific academic literacy intervention at the Potchefstroom Campus of the North-West University. The article addresses two research questions: 1) Has this specific academic literacy intervention had a significant impact on students' academic literacy levels?; and 2) Does using a variety of instruments lead to a more effective assessment of students' academic literacy levels?

## 2. Literature review

Academic literacy interventions are commonplace in most South African universities (Sebolai, 2014: 52; Fouché, 2009: 41-42), largely due to inadequate secondary education

in the country (Cliff, 2014: 322; Davies, 2010: xi) and the trend towards massification of tertiary education (Calderon, 2012; Teichler, 1998). However, few studies provide evidence of the impact of these interventions (Sebolai, 2014: 52; Butler, 2013: 80; Terraschke & Wahid, 2011: 174; Carstens & Fletcher, 2009: 319; Storch & Tapper, 2009: 218; Holder, Jones, Robinson & Krass, 1999: 20). There are two reasons why it is important to evaluate the effectiveness of interventions: firstly, it is important to be accountable to the various stakeholders in an intervention, and secondly, interventions should constantly be improved, which can only be done effectively if one knows which aspects of an intervention are effective, and which are not (Hatry & Newcomer, 2010: 678).

An obstacle to effective impact measurement of academic literacy interventions, specifically in South Africa, is that often, traditional experimental designs are not possible as all students partake in academic literacy interventions, making the use of control groups difficult or even impossible. One way of compensating for a lack of control groups is to conduct a quasi-experimental study in which data are triangulated (Lynch, 1996: 73-74). Triangulation is always preferable in any research study (Lynch 1996: 59-61, 2003: 152), but it becomes indispensable in validly and reliably assessing academic literacy interventions when traditional experimental designs cannot be utilised.

One common way of assessing students' academic literacy levels is by using a widely-used language proficiency test which has been proven to be valid and reliable. Studies that have done so include those by Petersen-Waughtal and Van Dyk (2011), Van Dyk, Cillié, Coetzee, Ross and Zybrands (2011), Wait (2007) and Song (2006). Such tests are used to firstly determine the effectiveness of an intervention (Henning (1987: 2) and to provide a backwash effect to the course developer to determine which curriculum outcomes have been successfully acquired by students, and which need to be approached differently in future (cf. Hughes 2003: 1-2). Another benefit of valid and reliable language proficiency tests is that one can account for a variety of external variables by analysing these tests statistically (for example by means of t-tests, correlations, and effect sizes).

Another frequently used method of measuring an improvement in students' academic literacy levels is by assessing student writing. Butler (2013: 83) argues that using additional assessment tools such as writing assessments is important, as relying only on multiple choice tests might not provide a sufficiently in-depth overview of students' academic literacy proficiency. Writing well is commonly associated with success not only academically, but also professionally (Weigle, 2002: 4). In fact, Weigle (2002: 5) argues that "[w]riting and critical thinking are seen as closely linked, and expertise in writing is seen as an indication that students have mastered the cognitive skills required for university work".

The literature provides several guidelines for creating appropriate writing assessments. The first guideline is that the topic should be well-defined and guide students in the writing process (Heaton 1988: 137, 144; Weigle 2002: 53, 93). Providing a single topic to assess students allows the evaluator to have a "common basis for comparison and evaluation" (Heaton 1988: 138; also see Shaw and Weir 2007: 247). Students should have a similar background knowledge of the topic, and the topic should not advantage

specific groups of students (Weigle 2002: 46). Furthermore, it might be useful to allow students to complete writing tasks at home to simulate real-life scenarios in which students have time to create drafts and edit their work multiple times – practices which are usually encouraged by academic literacy interventions (Heaton, 1988: 138). The reliability and construct validity of the scoring of writing assessments (cf. Bachman & Palmer, 1996: 17) must also be considered – both of these can be improved by using marking rubrics, as these allow the marker to separate various features of writing, and provide the marker with parameters of what is expected of students for various mark categories (Weigle, 2002: 51; Heaton, 1988: 148). Moreover, writing tasks should be authentic (cf. Bachman & Palmer, 1996: 23), and resemble those that students would be expected to complete in the academic literacy intervention, as well as in their content subjects. Finally, it is important to keep practicality in mind; that is, the resources that are available (for the administration and the assessment of the test) must be considered when setting up writing assessments, as these tend to be marking-intensive (Weigle, 2002: 54; Bachman & Palmer, 1996: 35).

Assessing students' academic literacy levels by means of a widely used valid and reliable academic literacy test, as well as by means of a writing assessment, are the two main instruments that are used for the current study.

### 3. Background

At the North-West University's Potchefstroom Campus, all first-year students are required to write the Test of Academic Literacy Levels (TALL) or its Afrikaans equivalent, the *Toets van Akademiese Geletterdheidsvlakke* (TAG)<sup>1</sup>. Students who are identified as 'at-risk' by the TALL (because they are not considered to have adequate academic literacy levels to successfully complete their university studies) are required to complete a semester-long academic literacy course. After completing this course, they join the students who were not identified as at-risk for a second semester-long academic literacy course (which is compulsory for all first-year students). Both of these courses are taught in the students' first year of studies.

The academic literacy intervention which is the focus of the current study is the first semester of a course called 'An introduction to academic literacy' (AGLE111). A total of 624 students registered for this course in 2015. The following outcomes were taken from the AGLE111 weekly lesson plan and serve to illustrate the nature of the course:

1. Identifying word meaning from context;
2. Paraphrasing text;

---

<sup>1</sup> The current study only focuses on the students who study through the medium of English, and will thus refer to only the TALL and the concomitant English academic literacy courses from this point onward.

3. Making effective notes from presentations and reworking these notes to paragraphs and mind maps;
4. Including references in a text;
5. Understanding academic genres and identifying and finding reliable academic sources;
6. Explaining the concepts of active reading, skimming and scanning;
7. Using skimming and scanning to obtain information from texts;
8. Identifying the qualities of, and being able to write good introductions and conclusions;
9. Creating a table of contents, and using it to plan and structure text;
10. Writing paragraphs with clear topic sentences, one main idea and applicable support;
11. Identifying action words and content words in examination questions and assignments, and planning well-structured responses to examination questions;
12. Identifying reasons for using the passive voice;
13. Identifying inaccurate information;
14. Writing correct sentences;
15. Calculating basic percentages;
16. Explaining and being able to identify visual manipulation;
17. Referring correctly to different parts of graphs and tables;
18. Identifying reasons for using graphic information, analysing graphics and discussing graphics appropriately;
19. Being aware of the structure of a seminar, being able to ask effective questions, and being able to answer questions effectively; and
20. Distinguishing between open, closed and hypothetical questions.

This generic academic literacy course, which is worth 12 credits, services all the faculties at the Potchefstroom Campus of the North-West University (thus, students belong to faculties as diverse as Arts, Natural Sciences, Theology, Education Sciences, Economic and Management Sciences, Law, Engineering, and Health Sciences). Students attend two one-hour classes per week which are embedded in the curriculum. The duration of the course is from February to May. The impact of AGLE111 has not yet been formally assessed. That, then, is the aim of Sections 4 and 5. The design and method of the study will be discussed in Section 4. The results of the study are thereafter discussed in Section 5.

#### **4. Design and method**

Four instruments (discussed below) were identified from the evaluation design proposed in Fouché et al. (2016) – see Appendix A for a summary of this evaluation design. These instruments were considered appropriate as they suit the generic nature of the subject (thus, the subject-specific instruments proposed in Fouché et al. [2016] would not be appropriate) and were considered feasible, taking into account the wide range of faculties and large number of students serviced by AGLE111. In addition, they could be easily integrated into the subject's existing assessment plan. The results of these instruments are triangulated by considering aspects of academic literacy that are clearly measured by both instruments, to determine whether results from the instruments are in line with each other, or possibly contradict each other. The results are analysed to determine whether there was an improvement in students' academic literacy levels between the onset and the completion of the AGLE111 academic literacy course. The instruments that were used are:

1. Using a generic academic literacy test;
2. Using a generic extended writing assignment (assessed by means of a rubric);
3. Quantitatively assessing an extended writing assignment; and
4. Correlating academic literacy achievements with students' results in their other subjects.

A sample of 173<sup>2</sup> (of the 624 registered students) was used for the first and the fourth instruments. A sample of 139 students was considered for the second instrument. Due to the labour intensity of the third instrument, it was not feasible to mark the same sample of 139 students that was used for the second instrument. A smaller random sample of 50 pre- and 50 post-assignments was thus used.

---

2 The maximum sample size that could be used for each instrument was used. Sample sizes are smaller than the total population as pre- and post-test data were not available for all students, and because many students did not give permission for their data to be used for research purposes.

The researcher had difficulty in obtaining electronic versions of students' post-writing assignments. Ultimately, a convenience sample was identified based on students who had submitted both pre- and post-writing assignments electronically. The pre-writing assignment results of the convenience sample were compared to the pre-assignment marks of a random sample to determine whether these groups could be considered comparable (and thus, whether the convenience sample could be said to be representative of the entire population, as would be the case with a random sample). A writing assessment rubric (see Section 5.2, Appendix B) was used to assess the writing of the two groups, and effect sizes were calculated for the difference in performance across the four main sections addressed in this rubric. A very small effect size (Cohen's  $d = 0.08$ ) was found between the results of the two groups, indicating that there was no practically significant difference between the groups. Thus, these two groups seem to be comparable, which means that it is likely that the convenience sample can be said to be representative of the entire AGLE111 population.

The following section reports on the results of the study by considering each of the four instruments selected for the current study.

## 5. Results and discussion

### 5.1 Instrument 1: Using a Generic Academic Literacy Test

Students wrote the 2015 version of the TALL in February as a pre-test, and the AGLE111 examination in June as a post-test. The TALL and the AGLE111 examination are not the same test. In this course's context, it was not practical to write the TALL as an examination as it does not sufficiently cover the abilities focused on in the AGLE111 course. Although the course (and its examination) is based on the TALL's theoretical construct, it does contain some elements not covered by the TALL, for example referencing abilities. A statistical item analysis could not be conducted to determine whether the TALL and the AGLE111 examination were statistically equivalent, since both tests were not written by the same population at the same point in time (Davies, Brown, Elder, Hill, Lumley & McNamara, 1999: 199).

Instead, a test-retest method with equivalent forms was decided upon for this study (cf. Miller, Linn & Gronlund, 2009: 110). According to Miller et al. (2009: 111), it is usually not "desirable to readminister the same assessment". This might be because of security reasons, or because students might remember answers from the previous assessment (Miller *et al.*, 2009: 398; Davies, Brown, Elder, Hill, Lumley & McNamara, 1999: 5, 198; Bachman, 1990: 182-183). Miller *et al.* (2009: 134) argue that comparable test forms "are especially useful in measuring development". The following aspects need to be considered when arguing for the equivalence of tests.

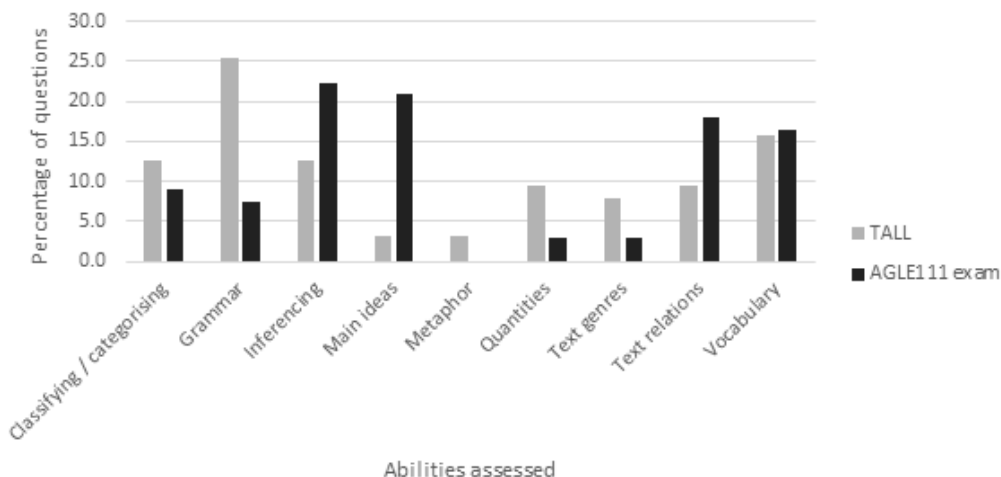
The reliability of the respective tests is the first aspect that should be considered. Using Chronbach's alpha (cf. Miller *et al.*, 2009: 135), the TALL (used as a pre-test in this study) had a reliability coefficient of 0.82 while the AGLE111 June examination had a reliability coefficient of 0.73 – both of these are considered strong reliability coefficients. Secondly, tests should be “designed to the same specifications” (Davies *et al.*, 1999: 5). This means that tests must measure the “same domain of achievement”, and that test items should be of a high technical quality, “developed by educational and test specialists” (Miller *et al.*, 2009: 399; see also Davies *et al.*, 1999: 198). Thirdly, test items should be similar in content, level of difficulty, number and type, have discriminating power, and be developed based on rigid and clearly defined specifications (Miller *et al.*, 2009: 134, 399; Davies *et al.* 1999: 5; Bachman, 1990: 183). Fourthly, the tests should have a score scale in common (Miller *et al.*, 2009: 134, 399). Finally, administration procedures, including instructions and length of tests, as well as scoring should be “so precisely stated that the procedures are standard for different users of the test” (Miller *et al.*, 2009: 399; see also Davies *et al.*, 1999: 5; Bachman 1990: 184). The two tests in the current study meet all of these criteria; therefore, it can be argued that the tests can be considered suitable for test-retest with equivalent forms.

Both assessments include the following sub-sections: understanding texts, vocabulary, grammar and text relations, scrambled text, and graphic and visual information. However, the first test has a 5-mark section called “text types” that is not included in the AGLE111 examination. Questions were analysed based on functions/abilities addressed by the questions in each test. The various headings are described more fully below:

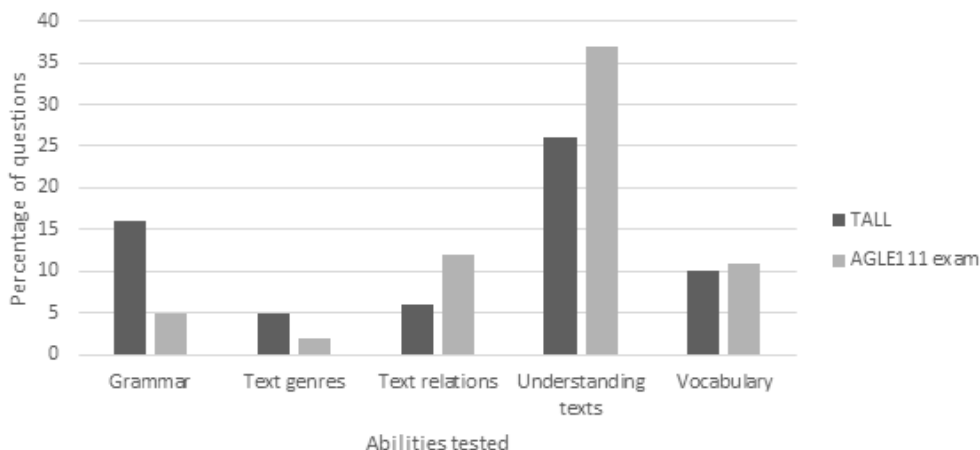
- **Classifying / categorising:** Classifying and categorising information
- **Grammar:** Using parts of speech and concord appropriately. Understanding sentence structure.
- **Inferencing:** Understanding and making inferences from information
- **Main ideas:** Identifying topics and main ideas; distinguishing between essential or non-essential information
- **Metaphor:** Using language metaphorically
- **Quantities:** Working with quantities / proportions / percentages without a calculator
- **Text genres:** Text types and text genres
- **Text relations:** Text relations and cohesion
- **Vocabulary:** Ability to make use of academic vocabulary in context

When questions are categorised based on the functions addressed by the questions in each test, the following is seen (see Figure 1). Both tests contain a similar percentage of questions (based on function type) that addresses vocabulary. There are fairly large discrepancies in the percentages of questions that address the remaining functions.





**Figure 1 Questions categorised based on abilities assessed**



**Figure 2 Questions categorised based on abilities assessed – re-categorised**

When the reliability of the two tests is examined, the following is found. The TALL boasts a high internal reliability measure (using Cronbach’s Alpha) of 0.82<sup>3</sup>, while the AGLE examination has a fair internal reliability rate of 0.73. Considering that both tests use a very similar theoretical construct, and that both have more than acceptable reliability measures (of more than 0.5; see Gliem & Gliem, 2003: 87), an improvement between pre- and post-test can be considered valid. However, most of the sub-sections in the two

<sup>3</sup> All TALL sub-sections have fair reliability rates of between 0.47 and 0.88 (using Cronbach’s Alpha) when outlier questions are removed.

tests cannot be similarly compared due to low internal reliability rates in three of the five AGLE sections, namely “Grammar and text relations”, “Scrambled text”, and “Graphic and visual information”, all of which have Cronbach Alpha scores of between 0.23 and 0.44. One reason for the low reliability rates for the “Grammar and text relations” as well as “Graphic and visual information” sections might be the order in which these were written in the two tests. In the TALL, the “Grammar and text relations” section appeared at the end of the test, whereas the “Graphic and visual information” section appeared at the end of the AGLE111 examination. The results indicated that several students did not complete the respective sections in the two tests. It follows that it is likely that some students were rushing to complete each test, which might affect the results in these two sections. Another possible reason for low reliability scores might be the number of questions addressed in each section. “Scrambled text”, for example, consisted of only five questions – such a low number of questions in a section often has a low reliability coefficient (Wells & Wollack, 2003: 5). Two sections which do, however, show valid internal reliability rates are “Understanding texts” and “Vocabulary”, with Cronbach Alpha scores of 0.59 and 0.54 respectively. One question in the “Vocabulary section” proved to be particularly problematic – if this question is removed, the Cronbach Alpha score improves from 0.54 to 0.60.

Both the TALL and the AGLE111 examinations were administered under similar conditions, using a multiple-choice format, and being written under examination conditions.

## Discussion

Students showed a slight improvement between writing the pre- and the post-test (see Table 1). The mean for the TALL was 46.16%, while the mean for the AGLE111 examination was 50.14, thus indicating an improvement of approximately 4% (Cohen’s  $d = 0.27$ ;  $p$ -value = 0.00). The medium effect size of 0.27 might be due to reliability concerns within the individual sections of the AGLE111 examination. Alternatively, it could be that the duration of the semester-long course was too short for a more practically significant improvement. When the improvement of the various sections that were present in both tests is considered, it would seem that students showed the greatest improvement in the questions relating to “Grammar and text relations”, “Understanding texts” as well as “Vocabulary”. However, due to the low internal reliability rate, the data obtained from “Grammar and text relations” will not be considered here. Nor will the data from the “Scrambled text” and “Graphic and visual information” sections be considered, for the same reason. Both of these sections showed a sharp decrease in marks (of approximately 10%) between the pre- and the post-test, which could be explained by the low reliability rates of these sections; thus, these sections seem not to have reliably measured the constructs they were meant to measure.

Of those sections that had an acceptable internal reliability rate, the greatest improvement could be seen in the “Understanding texts” section, in which students improved by 9.37%. The “Vocabulary” section also showed a big improvement of 8.46%. A paired

t-test was done to determine the significance of these improvements. Though both of these sections were statistically significant (p-value < 0.1), the effect sizes were small in both cases (0.21 and 0.37 respectively).

**Table 1: Improvement between the pre-intervention and post-intervention academic literacy tests (1<sup>st</sup> semester)**

	Mean TALL (%)	Mean AGLE111 examination (%)	Improvement (%)	Standard deviation (pre-test; post-test)	Sig. (2 tailed) (p-value)	Effect size **
<b>TOTAL</b>	46.16	50.14	3.98	14.95; 11.96	0.00*	0.27
<b>Understanding texts</b>	45.35	54.72	9.37	20.91; 14.72	0.00*	0.21
<b>Vocabulary</b>	43.99	52.45	8.46	17.61; 20.15	0.00*	0.19
<b>Grammar and text relations</b>	35.04	47.95	12.91	27.56; 12.65	0.00*	0.37
<b>Scrambled text</b>	67.86	49.13	-18.73	33.66; 26.28	0.00*	-0.28
<b>Graphic and visual information</b>	51.52	43.38	-8.14	21.79; 15.02	0.00*	-0.16

\* note: significant at p ≥ .001  
 \*\* Effect sizes: <.10: trivial; .10 - .30: small to medium; .30 - .50: medium to large; >.50: large to very large

A clear limitation to this instrument is that the same test was not used as both pre- and post-test, and that the AGLE111 examination had not been tested as valid and reliable by piloting it before students wrote it. However, in most contexts it would be impossible to do so for an examination that needs to be created anew each year. The validity of a test is usually only determined after the test has been written, using multiple sources of evidence. This is not practically possible in the case of most examinations; nor is piloting such an examination beforehand. In this context, using the TALL (which had been proven to be valid and reliable) as both pre- and post-test was not viable (as it did not cover all of the outcomes of the module), but future studies should aim to use existing valid and reliable tests as both pre- and post-tests. Where this is not possible, the researcher must keep in mind these limitations by using academic literacy tests in a pre-test / post-test scenario.

## 5.2 Instrument 2: Using a Generic Extended Writing Assignment (Assessed by Means of a Rubric)

Students completed a writing assignment at the beginning of the semester, and one at the end of the semester. These assignments were similar in the following aspects, as recommended in Section 2. Both were approximately two pages in length. Students were given a single topic as a pre-test topic, and a single topic as a post-test topic (Shaw & Weir, 2007: 247; Heaton, 1988: 138). Both pre- and post-assignment topics were of a generic nature, falling within the genre of the argumentative essay (thus, students had to argue a specific point), and were clearly defined in terms of the scope and content that had to be covered (see Weigle, 2002: 53, 93). Students were made aware of the audience for which they were writing for both of these assignments (see Weigle, 2002: 46). Furthermore, students' educational background, age and gender were taken into account in the choice of these topics (see Weigle, 2002: 46). Both topics required students to conduct some research, and called for similar background knowledge (see Hughes, 2003: 90): the first topic concerned the reasons for teenage smoking, and the second topic concerned gender equality. Thus, students would be able to respond to both topics using the level of background knowledge that a first-year student is likely to have, in addition to some literature to support their arguments. Students had time to complete both writing assignments at home, giving them enough time to go through necessary drafting processes and to conduct necessary research (see Weigle, 2002: 52; Heaton, 1988: 138).

The main difference between the two assignments was that some structure was provided in the first assignment. An introductory paragraph as well as four further introductory phrases were provided in this assignment, whilst no similar structure was provided in the post-assignment. This scaffolding provided in the first assignment was not considered necessary in the second as students had been taught to provide this structure in the AGL111 course they had just completed.

Both sets of assignments were assessed by two marking assistants. A marking session was held at the beginning of the process to ensure that marking assistants marked at the same level. Furthermore, the first five assignments marked by each marking assistant were sent back to the researcher and returned with detailed feedback. Thereafter, the researcher moderated a sample of all marked assignments to ensure that the marking level stayed consistent throughout the marking process. To further limit inconsistencies in marking, the same rubric (based on Carstens & Fletcher, 2009: 59-60) was used for both pre- and post-assignments (see Weigle, 2002: 22; Heaton, 1988: 148).

The original rubric was adapted after asking various academic literacy specialists to comment on it, and to recommend additional areas that were not yet covered by the rubric. The adapted rubric can be found in Appendix B. For the purposes of the current study, some sub-categories from that rubric were left out as they were not relevant to this marking assignment. These headings were "Introduction" (as an introduction was provided in the pre-assignment, and no comparison could be drawn), "Thesis

statement” (for the same reason), “Technical vocabulary” (as this was a generic topic), and “Integration of visual data” (as no visual data were required by either of these topics). A decision was taken to approach “Paraphrasing information from source texts” differently – rather than giving students a mark out of seven for this section, all student assignments were analysed by means of Turnitin (LLC, 2015) to obtain a percentage that indicated how much plagiarism (if any) was committed. All texts were cleaned for this purpose by deleting original instructions (including the provided introduction and phrases of the pre-assignment), direct quotations as well as reference lists. Only five or more words in a row were considered plagiarism for the purposes of this assessment. This guideline was decided upon in consultation with the University’s Turnitin liaison, as a phrase length of five words or more was deemed suitable to identify actual plagiarism rather than common word clusters used by students.

The sections of the rubric have a high internal reliability. The various sections have the following Cronbach’s Alpha scores: “Structure and development” = 0.91; “Academic writing style” = 0.86; “Editing” = 0.66; and “Use of source material” = 0.85. With the exception of the “Paraphrasing information from source texts” section (which is indicated as a plagiarism percentage), students received a score out of seven. The following qualifiers were given for each of the seven options: 7 = Excellent; 6 = Very good; 5 = Good 4 = Average; 3 = Below average 2 = Poor 1 = Very poor.

## 7. Discussion

As can be seen in Table 2, students improved in all areas covered by the marking rubric. The sub-categories that showed the greatest improvement were those related to the use of source material, namely “Referencing technique”, “Relevance of source data”, “Appropriately citing quotations” and “Integration of source data with text”. It is however notable that more plagiarism was found in the post-test than in the pre-test (11.9% and 9% respectively, though this difference was not found to be statistically significant, with a p-value of 0.14).

A paired t-test was done to determine whether there were practically significant improvements in the various categories and sub-categories on which students were assessed. The improvement in all of the categories and sub-categories (with the exception of “Paraphrasing information from source texts”) was statistically significant (thus,  $p < 0.05$ ). If Cohen’s (1988) guidelines (of considering an effect size of 0.2 small, an effect size of 0.5 medium and an effect size of 0.8 large) are followed, then all of the categories and sub-categories were close to having either a medium or large effect size, with the greatest effect sizes being evident in the “Use of source material”, “Academic writing style” and “Editing” categories.

Students are usually not required to reference source material at secondary school level. It therefore makes sense that such a great improvement would be seen in the “Use of source material” section between the pre-assignment (at which time students had

received little or no instruction on this topic) and post-assignment (at which time students had received instruction on this topic).

In addition to the use of source material in texts, the academic literacy course in question also seems to have impacted all other categories on this rubric, in particular “Editing” and “Academic Writing Style” (both of which showed big improvements as well as large effect sizes of 0.85 and 0.90 respectively) – all of which form part of the English syllabus at secondary school level (as opposed to referencing source material, as discussed above). Thus, based on the results in Table 2, the academic literacy course does seem to be fulfilling its purpose in addressing the poor academic literacy levels with which students enter universities.

**Table 2: Improvement between the pre- and post-intervention writing assignments (assessed by means of rubric) (1<sup>st</sup> semester)**

	Mean pre-assignment / 7	Mean post-assignment / 7	Improvement / 7	Standard deviation (pre-assignment; post-assignment)	Sig. (2 tailed) (p-value)	Effect size **
<b>STRUCTURE AND DEVELOPMENT</b>	<b>4.33</b>	<b>4.79</b>	<b>0.46</b>	<b>0.88; 0.75</b>	<b>.000*</b>	<b>0.53</b>
1. Development of main argument	4.4	4.8	0.4	0.99; 0.98	.001*	0.39
2. Paragraph development (topic sentences, main ideas, supporting information)	4.6	5	0.4	1.00; 0.93	.000*	0.41
3. Relevance of content to topic	4.7	5.3	0.6	1.05; 0.86	.000*	0.64
4. Linking devices (structuring thought with discourse markers, pronouns etc)	3.9	4.4	0.5	1.01; 0.91	.000*	0.43
5. Conclusion	4.1	4.6	0.5	0.99; 1.09	.000*	0.41
<b>ACADEMIC WRITING STYLE</b>	<b>3.95</b>	<b>4.75</b>	<b>0.79</b>	<b>0.88; 0.70</b>	<b>.000*</b>	<b>0.90</b>
6. Syntax: phrase and clause structure	3.6	4.2	0.6	1.03; 0.97	.000*	0.60
7. Academic vocabulary	3.9	4.8	0.9	1.08; 0.93	.000*	0.81
8. Style (formality; rhetorical mode)	4.4	5.2	0.8	1.03; 0.78	.000*	0.86

<b>EDITING</b>	<b>3.99</b>	<b>4.69</b>	<b>0.70</b>	0.82; 0.76	.000*	0.85
9. Spelling, capitalisation and punctuation	3.8	4.3	0.5	1.17; 1.02	.000*	0.38
10. Concord and tense	4.1	4.9	0.8	0.94; 0.95	.000*	0.95
11. Layout and format	4.1	4.9	0.8	0.98; 1.10	.000*	0.79
<b>USE OF SOURCE MATERIAL</b>	<b>2.91</b>	<b>4.53</b>	<b>1.62</b>	0.87; 0.94	.000*	1.86
12. Referencing technique	2.5	4.1	1.6	1.08; 1.16	.000*	1.45
13. Appropriately citing quotations	2.6	4.5	1.9	0.92; 1.22	.000*	2.07
14. Relevance of source data	4.0	5.2	1.2	1.20; 0.99	.000*	1.00
15. Integration of source data with text (synthesising)	2.5	4.4	1.9	1.00; 1.25	.000*	1.82
<b>PLAGIARISM</b>						
16. Paraphrasing information from source texts (expressed as a plagiarism percentage)	9%	11.9%	-(2.4%)	17.12; 16.58	.140	0.14
* note: significant at $p \geq .001$						
** Effect sizes: <.10: trivial; .10 - .30: small to medium; .30 - .50: medium to large; >.50: large to very large						

The section with the lowest level of improvement (of 0.46 marks, or approximately 7%), namely “Structure and Development”, still shows a clear statistically significant improvement with a medium effect size. It is possible that the abilities addressed in this section (namely “Development of main argument”, “Paragraph development”, “Relevance of content to topic”, “Linking devices” and “Conclusion”) are complex and thus difficult to acquire – an even greater improvement might be evident after another semester of academic literacy study.

### **5.3 Instrument 3: Quantitatively Assessing an Extended Writing Assignment**

Students' writing assignments were further analysed by means of measures suggested by Storch and Tapper (2009) – this analysis is summarised in Appendix C. Storch and Tapper (2009) use the following measures: fluency, accuracy, use of academic vocabulary, as well as text structure and rhetorical quality. This study chose to only use two of these measures, namely accuracy and use of academic vocabulary. The sections below indicate why two measures, namely fluency and text structure, and rhetorical quality, were left out, and how the remaining two measures, namely accuracy and academic vocabulary, were measured.

#### ***Fluency***

Storch and Tapper (2009: 211) measured fluency by looking at the number of words written as well as the words per T-unit. For the purposes of this study, considering students' total number of words was not thought useful as students were given guidelines regarding length, and as a result their pre- and post-assignments were of similar length (on average 537 and 568 words respectively). Furthermore, considering words per T-unit was also not considered sensible for this study. In the case of the Storch and Tapper (2009) study, the authors believed that a longer T-unit would equate to more fluent writing, which may well be the case for postgraduate students. However, in the case of first-year students, the aim is rather to encourage students to write shorter sentences (which may or may not be reflected in shorter T-units). Thus, T-unit length was not considered suitable for this study. It should be noted that even in the Storch and Tapper (2009) study, no statistically significant difference could be found between pre- and post-scores with regards to words per T-unit. It is possible that fluency is best measured using a more traditional rubric, and that a quantitative approach to measuring fluency is not necessarily effective for the purpose of most academic writing. An alternative option that was not explored for the purposes of the current article would be to do frequency counts of words that are typically used to enhance the fluency of a text (such as conjunctions or pronouns), to determine whether more such words are used after the conclusion of an intervention than was the case before the commencement of the intervention.

#### ***Rhetorical quality***

Storch and Tapper (2009: 212) assessed rhetorical quality by considering the overall structure, coherence and cohesion, and content of students' writing. A traditional marking rubric (Appendix B) was used for this purpose. The subsections on this marking rubric were very similar to those under the "Structure and development" section used in the current study (discussed in Section 5.2). It was thus considered unnecessary to replicate this assessment using an instrument that was so similar to the one already used.



## **Accuracy**

Storch and Tapper (2009: 219-220) coded assignments using 18 categories of errors. For the current study, a decision was taken to merge some of these categories to ultimately use only ten categories of errors. After consulting with various colleagues, it was decided that some of the distinctions seemed unnecessary for first-year students (for example, it was considered sufficient to categorise word choice, collocation and derivational errors under the same heading, namely “word choice”). In addition, decreasing the number of categories reduced the complexity of the marking, and facilitated increased conceptual clarity, as many of the previous categories were difficult to distinguish from each other in practice. Figure 3 indicates how categories were merged.

To compensate for the slight difference in structure between the pre- and post-assignments (as discussed earlier), the paragraph and phrases that were provided in the pre-assignment were deleted from the assignments assessed in this sample. Furthermore, students’ plagiarism counts were subtracted from the total word count, as sections plagiarised from the Internet and academic articles were unlikely to contain any mistakes. The word count of assignments was thus adapted to only reflect the words that the students wrote themselves. Raw error counts were converted to reflect an error count per 100 words so as to compensate for the different lengths of students’ assignments.

**Figure 3: Categories used to calculate accuracy in written assignments**

Current study	Storch and Tapper (2009) study
Article / determiner errors	Determiners Articles
Plural errors	Plural
Word choice errors	Word choice Collocation Derivational (word form) errors
Agreement errors	Agreement (noun or pronoun with verb) Agreement of verb with subject.
Tense errors	Tense and verb form
Preposition errors (note: unlike in the Storch and Tapper [2009] study, incorrect prepositions were coded as preposition errors, and not as word choice errors)	Prepositions
Syntax errors	Word order Absence of major constituent Absence of minor constituent Passive form
Linking errors	Errors in linking ideas
Possessive errors	Possessive
Punctuation errors (note: unlike in the Storch and Tapper [2009] study, all punctuation errors were counted, and not just those that obstructed meaning)	Punctuation Capitalisation

Due to the high frequency of mistakes in the pre- and post-assignments of the current study (almost double the mistakes that were recorded in the Storch and Tapper [2009] study), it was decided not to calculate a ratio of error free T-units per total T-units or error free clauses per total clauses as was done in the Storch and Tapper [2009] study. The main concern with these measurements is that these ratio scores cannot distinguish between T-units and clauses which have only one error, and those with more than one error (see, for example, Storch & Tapper, 2009: 211; Bardovi-Harlig & Bofman, 1989: 22). Rather, a measurement of errors per 100 words (the third measurement used by Storch and Tapper [2009]) was considered a more useful representation of accuracy.

## **Academic vocabulary**

Storch and Tapper (2009: 212) assessed students' use of academic vocabulary by using Coxhead's (2000) Academic Word List (AWL). They counted a) the number of AWL types, b) the number of AWL tokens and c) the numbers of tokens as a percentage of words written. The current study used the same approach, but also added the number of AWL word families. Further, in addition to the number of tokens, the number of AWL families and types were also calculated as a count per 100 words (thus, as a percentage of words written) to compensate for different essay lengths.

A word family is a collection of variations on a word. For example, the words "hypothesis", "hypotheses", and "hypothesise" would belong to one word family. The AWL contains 570 such word families – this accounts for roughly ten percent of all words in academic texts (while these 570 families only account for 1.4% of words in fiction) (Coxhead, 2000: 213). A word type count indicates how many unique words from the AWL were used. Thus, if the student used the word "hypothesis" twenty times, this would only count as one word type, whereas it would count as twenty AWL tokens.

Cobb's (2015) Vocabprofile, which was adapted from Heatley, Nation and Coxhead (2002), was used for all of these word frequency counts. As was the case in Section 5.2 and the "Accuracy" sub-section of this section, all assignments were cleaned by deleting any provided paragraphs, phrases and instructions. Plagiarism was compensated for by deleting any plagiarised sections.

## Discussion

### **Accuracy**

Although all error categories showed a reduction in mistakes between the pre- and the post-assignment, very few showed a statistically significant reduction in errors (see Table 3). There was a substantial reduction in total errors per 100 words (this number dropped by almost half), but the reduction is not significant due to the large standard deviation (36.32) in the pre-test scores. The categories that did show a statistically significant reduction in errors are "Article / determiner errors", "Plural errors", "Preposition errors" and "Syntax errors". All of these categories also show medium effect sizes. This information corresponds to the related categories discussed in Section 5.2, namely "Spelling, capitalisation and punctuation", "concord and tense" as well as "syntax: phrase and clause structure" that also showed only medium effect sizes and moderate improvement. It is interesting to note that a large improvement was seen in a related category on the TALL/AGLE pre- and post-tests, in contrast to the average improvement observed using the accuracy score in the current section. The "Grammar and text relations" section on these tests showed a 12.91% (Cohen's  $d = 0.37$ ;  $p$ -value = 0.00) improvement.

**Table 3: Reduction of language mistakes between pre- and post-assignments (1<sup>st</sup> semester)**

	Mean pre-assignment (errors per 100 words)	Mean post-assignment (errors per 100 words)	Reduction in mistakes between pre- and post-assignments	Standard deviation (pre-assignment; post-assignment)	Sig. (2 tailed) (p-value)	Effect size ***
<b>Total errors</b>	16.93	8.48	8.45	36.32; 3.47	.097	-0.23
<b>Article / determiner errors</b>	0.82	0.33	0.49	1.71; 0.34	.039*	-0.28
<b>Plural errors</b>	0.61	0.28	0.34	0.87; 0.35	.008**	-0.39
<b>Word choice errors</b>	1.7	1.48	0.21	1.58; 0.77	.334	-0.13
<b>Agreement errors</b>	0.87	0.38	0.49	1.89; 0.35	.085	-0.26
<b>Tense errors</b>	0.32	0.18	0.14	0.46; 0.21	.086	-0.30
<b>Preposition errors</b>	0.72	0.4	0.32	0.76; 0.35	.010**	-0.43
<b>Syntax errors</b>	2.54	1.86	0.68	2.47; 0.86	.036*	-0.28
<b>Linking errors</b>	0.32	0.26	0.06	0.45; 0.26	.399	-0.13
<b>Possessive errors</b>	0.29	0.11	0.17	0.79; 0.19	.799	-0.22
<b>Punctuation</b>	3.27	2.03	1.24	6.24; 1.23	.142	-0.20

\* note: significant at  $p \geq .05$

\*\* Significant at  $p \geq .01$

\*\*\* Effect sizes: <.10: trivial; .10 - .30: small to medium; .30 - .50: medium to large; >.50: large to very large

### **Academic vocabulary**

Students used almost double as many words from the AWL in the post-assignment when compared to the pre-assignment (see “Tokens per 100 words”). As is evident from Table 4, this improvement is reflected in the increased number of word families per 100 words and word types per 100 words, all of which increased significantly with medium to large

effect sizes. The most practically significant indicator of increased academic vocabulary usage is the tokens per 100 words. The most useful indicators might, however, be AWL families and AWL types, as these two indicators demonstrate how many new academic words each student used in the post-assignment in contrast to the pre-assignment. A limitation of this instrument is that it merely measures the usage frequency of words on the AWL, and does not measure whether students used these words correctly. This limitation can be (and was) addressed by considering feedback received by means of other instruments, such as the writing rubric (see Section 5.2).

**Table 4: Increase in academic vocabulary usage between pre- and post-assignments (1<sup>st</sup> semester)**

	Mean pre-assignment	Mean post-assignment	Improvement between pre- and post-assignments	Standard deviation (pre-assignment; post-assignment)	Sig. (2 tailed) (p-value)	Effect size **
<b>Academic Word List Families per 100 words</b>	5.3	7.3	1.9	6.10; 4.65	0.008*	0.26
<b>Academic Word List Types per 100 words</b>	5.9	7.6	1.8	7.10; 4.57	0.008*	0.24
<b>Academic Word List Tokens per 100 words</b>	8.3	13.3	5.0	10.44; 6.70	0.000*	0.45

\* Significant at  $p \geq .01$

\*\* Effect sizes: <.10: trivial; .10 - .30: small to medium; .30 - .50: medium to large; >.50: large to very large

The results obtained from counting the frequency of AWL words in students' writing correspond to those obtained from the rubric used in Section 5.2. Results from the rubric indicate that there was a practically significant improvement of approximately 13% (Cohen's  $d = 0.81$ ;  $p$ -value = 0.00) in students' use of academic vocabulary. Another perspective on students' vocabulary use can be found by looking at the accuracy scores obtained from using Storch and Tapper's (2009) error categories. One of these categories is "word choice errors". This category does not consider increased academic vocabulary usage, but rather looks at whether students use (all) words correctly (i.e. accuracy). There was a small decrease in students' word choice errors, though this was not statistically (neither theoretically nor practically) significant. Even though students might have used

more academic words in the post-assignment than in the pre-assignment, the level of accuracy of their vocabulary usage therefore seems to remain largely unchanged.

#### 5.4 Instrument 4: Correlating Academic Literacy Achievements with Other Variables

A final instrument that was used for the current study was correlating students' academic literacy achievements with other variables. Firstly, students' AGLE111 examination results were correlated with their final AGLE111 marks. Secondly, students' AGLE111 examination marks, their AGLE111 final marks as well as their TALL marks were correlated with their results in their other subjects.

#### Discussion

Non-parametric correlations were drawn using Spearman's rho (see Table 5). There was a significant and strong correlation ( $r_s = 0.65$ ;  $p\text{-value} = 0.00$ ) between students' AGLE111 examination marks and their final AGLE111 marks, indicating that the module's continuous assessment was well in line with the examination. There was also a significant and strong correlation ( $r_s = 0.63$ ;  $p\text{-value} = 0.00$ ) between the TALL and the AGLE111 examination, further strengthening the assumption that these tests (in their entirety, though not necessarily all of their separate sections) can be considered equivalent.

**Table 5: Non-parametric correlations using Spearman's rho (1<sup>st</sup> semester)**

		AGLE111 exam	TALL	AGLE module	Marks in all other subjects
<b>AGLE111 exam</b>	Correlation Coefficient ( $r_s$ )	1.000	.628*	.649*	.237*
	Sig. (2-tailed)		.000	.000	.003
	N	173	169	161	161
<b>TALL</b>	Correlation Coefficient ( $r_s$ )	.628*	1.000	.508*	.211*
	Sig. (2-tailed)	.000		.000	.008
	N	169	169	157	157
<b>AGLE module</b>	Correlation Coefficient ( $r_s$ )	.649*	.508*	1.000	.347*
	Sig. (2-tailed)	.000	.000		.000
	N	161	157	161	161
<b>Marks in all other subjects</b>	Correlation Coefficient ( $r_s$ )	.237*	.211*	.347*	1.000
	Sig. (2-tailed)	.003	.008	.000	
	N	161	157	161	161

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

Unlike previous studies (e.g. Van Dyk, 2015: 180; Van Rooy & Coetzee-Van Rooy, 2015: 10-11; Mhlongo, 2014: 80-82) which found strong correlations between the results of academic literacy modules and students' results in their other subjects, the current study found only a moderate correlation (of 0.35) between the final AGLE111 mark and students' marks in their other subjects. Thus, whereas one could infer from the studies mentioned above that better performance in the academic literacy interventions under discussion in these studies might lead to improved performance in students' other subjects, this study cannot safely make the same assumption. It must be kept in mind, however, that students had only completed one academic literacy semester at this point, in contrast to the studies mentioned above. A stronger correlation might be found after students have completed the second AGLE module (AGLE121). A further fact which could influence this correlation is that the current study only focuses on those students who were identified as at-risk at the beginning of the year. The weak correlation (of 0.21) between the TALL and students' marks in other subjects is to be expected, as one would assume that if the academic literacy intervention had any impact, this impact would weaken any stronger correlation that would have existed had no intervention occurred.

## 6. Summary of findings

Students seem to have improved across a wide spectrum of academic literacy abilities after taking a one-semester long (from February to May) academic literacy course. Some notable improvements between the beginning and the end of the semester are students' ability to use source material in written assignments, an increase in students' academic vocabulary range and usage, students' ability to edit their work (including a reduction in language errors as well as an improvement in the layout and format of their assignments) and their ability to structure and develop their writing.

From the data obtained in this study, it may be inferred that the course has successfully addressed the following outcomes (see Section 3 for a complete list of the course's outcomes):

- Identifying word meaning from context (see Section 5.1, "Vocabulary" and "Understanding texts"; Section 5.2, "Academic vocabulary"; and Section 5.3, "Word choice errors" in the "Accuracy" sub-section as well as word frequency counts in the "Vocabulary" sub-section);
- Including references in a text (see Section 5.2, "Use of source material");
- Understanding academic genres and identifying and finding reliable academic sources (see Section 5.2, "Relevance of source data");
- Identifying the qualities of, and being able to write good introductions and conclusions (see the "Conclusion" category in Section 5.2);

- Creating a table of contents, and using it to plan and structure text (see Section 5.2, “Layout and format”);
- Writing paragraphs with clear topic sentences, one main idea and applicable support (see Section 5.2, “Paragraph development”); and
- Writing correct sentences (see Section 5.1, “Grammar and text relations”; Section 5.2, “Syntax”, “Spelling, capitalisation and punctuation”; and “Concord and tense” as well as the accuracy section of Section 5.3).

The performance in some outcomes did not seem to have improved between the beginning and the end of the semester. One such outcome was “paraphrasing text”. It should however be kept in mind that the only measure of students’ ability to paraphrase in the instruments used for this study was students’ plagiarism counts, which increased slightly (though not statistically significantly) between the beginning and the end of the semester. It is possible that the way in which they paraphrased information had indeed improved – students’ improved accuracy counts (Section 5.3) as well as improved “Syntax”, “Spelling, capitalisation and punctuation” as well as “Concord and tense” scores might be an indication that paraphrased sections were dealt with more effectively. Another outcome where students seemed to have performed more poorly in the post-test than in the pre-test is “Identify reasons for using graphic information, analyse graphics and discuss graphics appropriately” (Section 5.1), although this might be explained by the low internal reliability of this section in the AGLE111 examination. More information could have been collected had this outcome also been assessed in the writing assignment.

Some outcomes could not be assessed by means of the instruments that were used. In some cases, that is because appropriate instruments were not available; for example, no appropriate instrument could be found with which to assess students’ study strategies or time management abilities. In other cases, it is because it is not necessarily possible to directly measure some outcomes. Outcomes where this might have been the case include “Using skimming and scanning to obtain information from texts”, “Identifying action words and content words in examination questions and assignments, and planning well-structured responses to examination questions”, and “Being aware of the structure of a seminar, being able to ask effective questions, and being able to answer questions effectively”. Several outcomes which were not directly measured were, however, implicitly measured by measuring other outcomes; for example, students’ ability to read effectively was indirectly measured in that they had to conduct research and understand this research so as to report on it in their written work. Similarly, students would presumably be able to identify the reasons for using the passive voice if this ability is used appropriately in their writing.

In some cases, outcomes were not assessed by means of the instruments that were used, but could easily have been assessed had the instruments been adapted slightly. For example, the writing assignment could have required students to refer “correctly to different parts of graphs and tables”. It is therefore advisable to first consider all course outcomes before setting up assignments to be used as assessment instruments (also



see the comment above about the “Identify reasons for using graphic information, analyse graphics and discuss graphics appropriately” outcome).

Strong correlations could be found between the TALL and the AGLE111 examination, as well as between the AGLE111 examination and the total AGLE111 module, indicating that the pre- and post-test, as well as the examination and the continuous assessment of the AGLE111 module are all well aligned. However, as only a moderate correlation could be found between the AGLE111 module mark and students’ other marks, no certain claims can be made indicating that the AGLE111 module on its own is likely to lead to improved marks in students’ other subjects. Such an effect, however, might be more apparent after the second semester, when students have been able to implement all the abilities acquired throughout the AGLE111 as well as the AGLE121 modules.

## 7. Conclusion

In answer to the first research question, namely whether this specific academic literacy intervention has significantly impacted on students’ academic literacy levels, the course seems to effectively address several academic literacy abilities. However, it could potentially be improved by focusing more on abilities that showed no, or only a moderate, improvement, for example students’ visual literacy abilities, their ability to paraphrase information, and even their ability to structure assignments appropriately (the moderate improvement in this ability might have been influenced by the scaffolded approach that was followed in the pre-assignment – see Section 5.2). It should be kept in mind though that these abilities do receive more attention in the second semester module (AGLE121), and a clear improvement might only be visible after the completion of that module.

As far as the evaluation of academic literacy interventions in general is concerned, in answer to the second research question, it would seem that using a variety of instruments is indeed worthwhile, as richer information is consequently provided, thus assisting the researcher in correctly identifying the intervention’s strengths and weaknesses. Using a variety of instruments made it possible to either corroborate data (for example considering the improvement in students’ language usage by using instruments 2 and 3) or to provide a richer context for deductions made from data. An example of this is when one considers students’ vocabulary usage. While it seems clear that students use a wider range of academic vocabulary, and use it more frequently, in the post-assignment than in the pre-assignment (taking into consideration instrument 2 and the vocabulary section of instrument 3), they still make a similar number of vocabulary errors after a semester-long academic literacy course.

Future research will assess the increase and possible transfer of academic literacy abilities at the end of a year of academic literacy study. Pre- and post-tests and assignments will once again be used, in addition to perceptual data from students. Once this has been done, it is hoped that a complete picture of the impact of a year of academic literacy study at a South African university will emerge.

# Appendix A: Proposed evaluation design for academic literacy intervention

Appendix A: Proposed evaluation design for academic literacy intervention

GENERIC ACADEMIC LITERACY INTERVENTIONS	SUBJECT-SPECIFIC ACADEMIC LITERACY INTERVENTIONS	COLLABORATIVE ACADEMIC LITERACY INTERVENTIONS	LIMITED PURPOSE INTERVENTIONS (e.g. writing centres or reading programmes)
<b>Recommended instruments</b>	<b>Recommended instruments</b>	<b>Recommended instruments</b>	<b>Recommended instruments</b>
Generic academic literacy test	Subject-specific extended writing assignment (assessed by means of a rubric)	Subject-specific extended writing assignment (assessed by means of a rubric)	Student questionnaire
Student questionnaire	Subject-specific academic literacy test	Subject-specific academic literacy test	<b>Additional optional instruments</b>
<b>Additional optional instruments</b>	Student questionnaire	Student questionnaire	Subject-specific academic literacy test
Generic extended writing assignment (assessed by means of a rubric)	<b>Additional optional instruments</b>	Qualitative feedback from primary stakeholders	Subject-specific extended writing assignment (assessed by means of a rubric)
Generic extended writing assignment (assessed by means of quantitative measures)	Generic academic literacy test	<b>Additional optional instruments</b>	Generic academic literacy test
Lecturer questionnaire	Generic extended writing assignment (assessed by means of quantitative measures)	Generic academic literacy test	Generic extended writing assignment (assessed by means of quantitative measures)
Qualitative feedback from primary stakeholders	Lecturer questionnaire	Generic extended writing assignment (assessed by means of quantitative measures)	Lecturer questionnaire
Content analysis of study material	Qualitative feedback from primary stakeholders	Lecturer questionnaire	Qualitative feedback from primary stakeholders
Correlating academic literacy achievements with other variables	Content analysis of study material	Content analysis of study material	Content analysis of study material
	Correlating academic literacy achievements with other variables	Correlating academic literacy achievements with other variables	Correlating academic literacy achievements with other variables

- Where control group studies are possible to strengthen the research design, these should be used in combination with at least two instruments above.
- Where control group studies are not possible, findings should be validated by means of triangulation by choosing at least three of the options above when evaluating any given academic literacy intervention.
- In both scenarios above, one of the instruments should ideally be an instrument that measures the improvement in academic literacy abilities (coloured blue), and at least one should be an instrument that measures transfer (coloured green).

## Appendix B: Revised writing rubric

	1 (Poor)	2 (Below average)	3 (Average)	4 (Good)	5 (Excellent)	N/A
<b>Structure and development</b>						
1. <b>Introduction</b>	No attempt at an introduction is evident.		An attempt at an introduction is evident.		The introduction effectively introduces the topic, grabs the reader's interest, and indicates what will be discussed in the text.	
2. <b>Thesis statement</b>	No thesis statement is evident.		A thesis statement is present, but it is not as clear as it could be, and/or not sufficiently proven by means of argument.		A clear and logical thesis statement is provided that is proven by means of argument.	
3. <b>Development of main argument</b>	No argument is made in this text.		The text contains an argument, though this is not always clear and logical.		A clear and logical argument is built throughout the text.	
4. <b>Paragraph development (topic sentences, main ideas, supporting information)</b>	Paragraphs do not have topic sentences. Paragraphs do not centre around one main idea. The information in the paragraph does not support the topic sentence.		Many paragraphs have clear topic sentences, centre around a single main idea, and contain supporting information that relates to the topic sentence.		All paragraphs have clear topic sentences, centre around a single main idea, and all supporting information relates to the topic sentence.	
5. <b>Relevance of content to topic</b>	Content is not relevant to the topic.		Content is mostly relevant to the topic.		Content is highly relevant to the topic.	
6. <b>Linking devices (structuring thought with discourse markers, pronouns etc.)</b>	No attempt is made at using linking devices effectively.		An attempt at using linking devices effectively is evident throughout the text.		Linking devices are used logically and appropriately throughout the text.	

7.	<b>Conclusion</b>	No attempt at a conclusion is evident.	An attempt at a conclusion which summarises the main points of the text, ties ideas together, and finishes off the text is evident.	The conclusion succinctly and effectively summarises the main points of the text, ties ideas together, and finishes off the text.
<b>Academic writing style</b>				
8.	<b>Syntax: phrase and clause structure</b>	Numerous incomplete sentences and incorrect clauses are evident.	Complete sentences with correctly formulated clauses are generally evident.	Complete sentences with correctly formulated clauses are used throughout.
9.	<b>Academic vocabulary</b>	No attempt is made to correctly make use of academic vocabulary.	Some attempt is made to correctly make use of a range of academic vocabulary.	A wide range of academic vocabulary is used correctly throughout the text.
10.	<b>Technical vocabulary</b>	No attempt is made to correctly make use of technical vocabulary.	Some attempt is made to correctly make use of a range of technical vocabulary.	A wide range of technical vocabulary is used correctly throughout the text.
11.	<b>Style (formality; rhetorical mode)</b>	The writing style does not reflect the level of formality of the text.	The writing style generally reflects the level of formality of the text.	The writing style excellently reflects the level of formality of the text.
12.	<b>Genre</b>	The writing style is not appropriate for the intended genre.	The writing style is partially appropriate for the intended genre.	The writing style suits the intended genre perfectly.
13.	<b>Audience</b>	The writing style is not appropriate for the intended audience.	The writing style is partially appropriate for the intended audience.	The writing style suits the intended audience perfectly.
14.	<b>Integration of visual data</b>	Visual data are not integrated effectively at all.	An attempt to integrate visual data into the text is clear.	Visual data are fully and effectively integrated with the text, using correct captions with clear references to the visual data throughout the text.

Editing					
15.	<b>Spelling, capitalisation and punctuation</b>	Spelling, capitalisation and punctuation errors obstruct the meaning of the text.	Spelling, capitalisation and punctuation are generally accurate.	Spelling, capitalisation and punctuation are accurate throughout the text.	
16.	<b>Concord and tense</b>	Numerous concord and tense errors are evident in the text.	Some concord and tense errors are evident in the text.	No concord and tense errors are evident in the text.	
17.	<b>Layout and format</b>	The format requirements have not been met.	Format requirements are generally adhered to.	Format requirements are strictly adhered to. The format is polished and professional.	
<b>Use of source material</b>					
18.	<b>Referencing technique (list of references)</b>	The referencing is absent or unsystematic, and the prescribed referencing system is not used.	The referencing is generally accurate using the prescribed referencing system.	The referencing is consistently accurate using the prescribed referencing system.	
19.	<b>Relevance of source data</b>	Source data are not relevant to the topic.	Source data are generally relevant to the topic, and indicate sufficient research.	Source data are very relevant to the text, and indicate extensive research.	
20.	<b>Appropriately citing direct and indirect quotations</b>	Quotations are not cited in the text.	Quotations are mostly cited in the text using the prescribed referencing system.	Quotations are consistently and accurately referenced in the text using the prescribed referencing system.	
21.	<b>Paraphrasing information from source texts</b>	Information from source texts is not paraphrased.	A clear attempt was made to paraphrase most of the information from source texts.	Information from source texts are paraphrased correctly throughout the text.	
22.	<b>Integration of source data with text (synthesising)</b>	No attempt to synthesise source data with the author's argument is evident.	A clear attempt to synthesise source data with the author's argument is evident.	Source data are excellently synthesised with the author's argument throughout the text.	

## Appendix C: Storch and Tapper quantitative analysis

Storch and Tapper (2009) use the following measures to quantitatively assess student writing<sup>4</sup>.

### A. Fluency

- Measured in terms of:
- Number of words
- Words per T-unit

### B. Accuracy

Scripts are coded for the following 18 categories of errors (categorised into six broad themes, namely “syntactical”, “morphological: nominal”, “morphological: verbal”, “grammatical”, “lexical” and “mechanics” (taken from Storch and Tapper 2009: 219-220)

#### Syntactical

1. Word order
2. Absence of major constituent, such as subject, verb, object
3. Absence of minor constituent (e.g. ‘Enterprises may not be professional [enough] to master the coordination of ...’)
4. Errors in linking ideas (missing, redundant, or incorrect)

#### Morphological:nominal

5. Plural
6. Agreement (noun or pronoun with verb)
7. Possessive

---

4 Text structure and rhetorical quality are also investigated in the Storch and Tapper study using a more traditional rubric. However, the current study is primarily interested in the quantitative measures employed in the Storch and Tapper study. As rubrics are dealt with separately in the current study, the section on text structure and rhetorical quality is not included in this summary.

### **Morphological:verbal**

8. Tense and verb form. Errors of tense, aspect, mood and form for the same verb were counted as one error.
9. Agreement of verb with subject. Agreement errors involving both subject and verb in the same phrase were counted as one error. e.g. 'Every details [detail] need [needs] to be considered'.
10. Passive form (missing or incorrect)
11. Derivational (word form) e.g. 'very technologic [technological] parameters'

### **Grammatical**

12. Determiners (e.g. this, that, it, those). Missing, redundant or incorrect. e.g. 'when building cantilever bridges. Those [These] bridges ...'
13. e.g. 'Looking at its [this] background and current situation, .'
14. Articles. Errors of article and noun plurals were counted as one error. e.g. When the context shows that ' the problem' should be 'problems', one error was counted.
15. Prepositions (missing or redundant)

### **Lexical**

16. Word choice. (Register errors such as 'lots of' were not included). e.g. 'Many countries still out of [lack] responsibility'. e.g. 'especially in developing countries, such as my hometown [home] – China'. Prepositions were coded as word choice if the choice was incorrect. e.g. 'The glaciers in [at] the two poles of the earth'.
17. Collocation. Erroneous expressions and phrasal verbs were counted as one error. e.g. the key of the [to] success e.g. I am interested to conduct [in conducting]. If meaning was so obscure that reformulation was impossible, a phrase or clause was counted as one collocation error. e.g. 'The definition should "with which" or "follow with" conclude the rights, the duties .' was one error.

### **Mechanics (Spelling omitted)**

18. Capitalisation
19. Punctuation (if meaning was affected)

A repeated error was counted each time it occurred. Errors were counted according to the minimal number of corrections required to make a phrase or clause error-free, while maintaining the apparent meaning indicated by the context. For example, when taking context into account, a minimum reformulation of the following sentence yields 5 errors.

The following accuracy scores were calculated:

- a ratio of error free T-units per total T-units (EFT/T),
- a ratio of error free clauses per total clauses (EFC/C),
- and the total number of errors per total number of words (E/W).

### **C. Use of academic vocabulary**

Measured by means of Academic Word List (AWL) developed by Coxhead (2000). Counts performed:

- Numbers of AWL types
- Numbers of AWL tokens
- Number of tokens as a percentage of words written



## REFERENCES

- Bachman, L.F.** 1990. *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L.F. & Palmer, A.S.** 1996. *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Bachman, L.F. & Palmer, A.S.** 2010. *Language assessment in practice*. Oxford: Oxford University Press.
- Bardovi-Harlig, K. & Bofman, T.** 1989. Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition* 11(1):17-34.
- Brown, J.D.** 2001. *Using surveys in language programs*. Cambridge: Cambridge University Press.
- Butler, G.** 2013. Discipline-specific versus generic academic literacy intervention for university education: An issue of impact? *Journal for Language Teaching* 47(2):71-87.
- Calderon, A.** 2012. *Massification continues to transform higher education*. Available: <http://www.universityworldnews.com/article.php?story>. [Accessed: 20 October 2013].
- Carstens, A. & Fletcher, L.** 2009. Quantitative evaluation of a subject-specific essay-writing intervention. *Southern African Linguistics and Applied Language Studies* 27(3):319-332.
- Cliff, A.** 2014. Entry-level students' reading abilities and what these abilities might mean for academic readiness. *Language Matters* 45(3):313-324.
- Cobb, T.** 2015. *Web Vocabprofile*. Available: <http://www.lexutor.ca/vp/eng/>. [Accessed: 5 March 2015].
- Coxhead, A.** 2000. A new academic wordlist. *TESOL Quarterly* 34(2):213-238.
- Davies, A.** 2010. Assessing the academic literacy of additional-language students. *Southern African Linguistics and Applied Language Studies* 27(3):xi-xii.
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T. & Mcnamara, T.** 1999. *Dictionary of language testing*. Cambridge: Cambridge University Press.
- Fouche, I.** 2015. Towards impact measurement: An overview of approaches for assessing the impact of academic literacy abilities. *Stellenbosch Papers in Linguistics* 44(1):19-35.

- Fouché, I.** 2009. *Improving the academic literacy levels of first-year Natural Sciences students by means of an academic literacy intervention*. Unpublished MA dissertation. Pretoria: University of Pretoria.
- Fouché, I., Van Dyk, T. & Butler, G.** 2016. Impact measurement: towards creating a flexible evaluation design for academic literacy interventions. *Stellenbosch Papers in Linguistics* 45(1):109-145.
- Gliem, J.A. & Gliem, R.R.** 2003. *Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales*. Columbus: Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education.
- Hatry, H.P. & Newcomer, K.E.** 2010. Pitfalls in evaluations. In J.S. Wholey, H.P. Hatry & K.E. Newcomer (Eds.), *Handbook of practical program evaluation*, 3<sup>rd</sup> ed.: 557-580. San Francisco: Jossey-Bass.
- Heatley, A., Nation, I.S.P. & Coxhead, A.** 2002. *RANGE and FREQUENCY programs*. Available: <http://www.victoria.ac.nz/lals/staff/paul-nation.aspx>. [Accessed: 20 October 2014].
- Heaton, J.B.** 1988. *Writing English language tests*. New York: Longman.
- Holder, G.M., Jones, J., Robinson, R.A. & Krass, I.** 1999. Academic literacy skills and progression rates amongst pharmacy students. *Higher Education Research & Development* 18(1):19-30.
- Hughes, A.** 2003. *Testing for language teachers*, 2<sup>nd</sup> ed. Cambridge: Cambridge University Press.
- LLC.** 2015. *Turnitin*. Available: <http://www.turnitin.com>. [Accessed: 2 February 2015].
- Lynch, B.K.** 1996. *Language program evaluation: Theory and practice*. Cambridge: Cambridge University Press.
- Lynch, B.K.** 2003. *Language assessment and programme evaluation*. Edinburgh: Edinburgh University Press.
- Mhlongo, G.J.** 2014. *The impact of an academic literacy intervention on the academic literacy levels of first year students: The NWU (Vaal Triangle Campus) experience*. Unpublished MA dissertation. North West University. Unpublished).
- Miller, M.D., Linn, R. & Gronlund, N.E.** 2009. *Measurement and assessment in teaching*, 10<sup>th</sup> ed. Upper Saddle River: Pearson.

- Newcomer, K.E. & Triplett, T.** 2010. Using surveys. In: Wholey, J.S., Hatry, H.P. & Newcomer, K.E. (eds). 3rd ed. San Francisco: Jossey-Bass.
- Petersen-Waughtal, M. & Van Dyk, T.** 2011. Towards informed decision making: The importance of baseline academic literacy assessment in promoting responsible university access and support. *Journal for Language Teaching* 45(1):99-114.
- Sebolai, K. 2014. Evaluating academic literacy teaching at a South African university : A case study of an academic literacy programme. *Journal for Language Teaching* 48(1):51-69.
- Shaw, S.D. & Weir, C.J.** 2007. *Examining writing: Research and practice in assessing second language writing*. Cambridge: Cambridge University Press.
- Song, B.** 2006. Content-based ESL instruction: Long-term effects and outcomes. *English for Specific Purposes* 25(4):420-437.
- Steyn, S.** 2012. *Investigating issues of equivalence in the design of parallel English and Afrikaans tests of academic literacy for grade 11 learners*. . University of the Free State. (Unpublished).
- Steyn, S.** Forthcoming. *Investigating issues of equivalence in the design of parallel English and Afrikaans tests of advanced language ability for learners at FET level*. North-West University. (Unpublished).
- Storch, N. & Tapper, J.** 2009. The impact of an EAP course on postgraduate writing. *Journal of English for Academic Purposes* 8(3):207-223.
- Teichler, U.** 1998. Massification: A challenge for institutions of higher education. *Tertiary Education & Management* 4(1):17-27.
- Terraschke, A. & Wahid, R.** 2011. The impact of EAP study on the academic experiences of international postgraduate students in Australia. *Journal of English for Academic Purposes* 10(3):173-182.
- Van Dyk, T.** 2015. Tried and tested. *Tijdschrift voor Taalbeheersing* 37(2):159-186.
- Van Dyk, T., Cillié, K., Coetzee, M., Ross, S. & Zybrands, H.** 2011. Ondersoek na die impak van 'n akademiese geletterdheidsintervensie op eerstejaargestudente se akademiese taalvermoë. *LitNet Akademies* 8(3):487-506.
- Van Dyk, T., Van Rensburg, A. & Marais, F.** 2011. Levelling the playing field: An investigation into the translation of academic literacy tests. *Journal for Language Teaching* 45(1):153-169.

- Van Rooy, B. & Coetzee-Van Rooy, S.** 2015. The language issue and academic performance at a South African university. *Southern African Linguistics and Applied Language Studies* 33(1):31-46.
- Wait, T.H.** 2007. *The relationship between the use of language learning strategies and performance on a standardised English proficiency test*. Unpublished Unpublished MA dissertation, Stellenbosch University, Stellenbosch.
- Weigle, S.** 2002. *Assessing writing*. Cambridge: Cambridge University Press.
- Wells, C.S. & Wollack, J.A.** 2003. *An instructor's guide to understanding test reliability*. Wisconsin: University of Wisconsin.
- Yeld, N.** 2010. Some challenges and responses: Higher education in South Africa. *Discourse* 38(1):24-36.

---

## ABOUT THE AUTHOR

### **Ilse Fouché**

University of Pretoria, Private bag X20, Hatfield, 0028  
Email: [ilse.fouche@up.ac.za](mailto:ilse.fouche@up.ac.za)

**Ilse Fouché** is an academic literacy lecturer in the University of Pretoria's Four-Year Programme. Her Doctoral study focused on impact evaluation in the field of academic literacy. Her research areas include the scholarship of teaching and learning in language teaching and academic literacy.