

# INFORMATIVE ASSESSMENT: A SUPPORTIVE TOOL FOR SYSTEMIC VALIDITY IN LANGUAGE EDUCATION

**C.L. du Plessis\***

e-mail: duplessiscl@ufs.ac.za

**C. Els\***

e-mail: ElsCA@ufs.ac.za

\*Department of English  
University of the Free State  
Bloemfontein, South Africa

## ABSTRACT

Reviewing academic programmes and curricula is a necessary part of ensuring theoretical accountability in respect of the education provided to university students. However, accountability is context dependent and requires scrutiny of multiple sources of information to inform course design, including the analysis and interpretation of assessment data. This latter part is easy to overlook, especially when the focus falls on curriculum completion. The aim of this article is to argue how the correlation, analysis and interpretation of different kinds of course content and data can be used to identify and support learning needs, and provide further justification for course design. In view of the reciprocity that exists between language teaching and testing, core principles used for test validation are also useful for validating course components. As illustration hereof, this article evaluates a tertiary language course offered to Education students by drawing on the notions of construct and content validity. Empirical data gleaned from a literacy survey and assessment artefacts are correlated and interpreted against course objectives and content. A need is identified for substantial revisions to the course in order to ensure closer alignment with students' learning needs and critical literacies. The study shows how accountability in language education coheres with the principle of systemic validity.

**Keywords:** language education, systemic validity, language assessment, course evaluation, validation, accountability, critical literacy awareness

## ACCOUNTABILITY IN LANGUAGE TEACHING AND TESTING

Endeavours to address low English language proficiency and academic literacy levels of students at South African tertiary institutions go back at least 18 years to the establishment of the Alternative Admissions Research Project (AARP) in 2000, and the development of tests such as the National Benchmark Tests (NBTs) and Test of Academic Literacy Levels (TALL)

(Du Plessis 2012). These tests have been subjected to much scrutiny and validated as external post-Grade 12 indicators of knowledge and ability for the purposes of placement in university programmes. However, attempts to address students' academic literacy and language proficiency needs at tertiary level are not always evaluated to the same extent. The training of pre-service educators is one case in question. Tertiary institutions are accused of producing teaching graduates with insufficient subject content knowledge and pedagogical ability, despite the literacy interventions. Not only are the newly qualified educators described as having low proficiency in the languages of teaching and learning, but their weak reading and writing abilities are said to undermine their teaching performance (CDE 2015; Grosser and Nel 2013; Nkosi 2015). The credibility of university language education programmes is thus being placed under suspicion together with the dubious matriculation results released annually by the Department of Basic Education.

In light of the alleged inadequate preparation of educators for meaningful classroom engagement, this article investigates the relevance of an English language development course prescribed for a group of Education students. The objective is to ascertain whether the module has been designed optimally to provide students with the language skills and literacies that they need to teach effectively in schools. The focus of the article, however, will not fall on second language acquisition (SLA) from the perspective of "knowledge internalization, knowledge modification, and knowledge consolidation" (De Graaff and Housen 2011, 731), in other words the psycholinguistic operationalization of language learning. Rather, the study seeks a justification for the course design. The authors share the view that language instruction does facilitate language learning (Norris and Ortega 2000; Ellis 2008), and that it constitutes an indispensable part of the training of educators.

At the university in question, all Education students offer a number of credits in English. When examining the usefulness of a prescribed course, it should be borne in mind that accountability depends on more than predetermined standards and curricular content. Weideman (2017a, 2) refers to the importance of "locally relevant and contextualized codes of practice", and the necessity of employing a design framework that supports the ethicality of what we do. In this sense the principle of ensuring that our courses have utility and accountable effects, applies just as much to the course design as to the assessment instruments used.

Validation studies are commonly undertaken to justify the use of standardised language tests,<sup>1</sup> but there is also a need to justify the usefulness of our language development modules and accept responsibility for the way in which we train prospective teachers. In view of the reciprocity that exists between language teaching and testing, there is much to be gained from applying fundamental principles used for the responsible design and validation of language tests

(Weideman 2017b) to the review of a language module. The appraisal process should be elucidated by inferences based on formative and summative assessment data in order to support synergy of instruction and assessment. Insights gained from course results and test data provide essential evidence either in support of the theoretical justification for the language course, or in refute thereof.

At this point, the notion of construct validity in language testing needs some elucidation. The modern orthodoxy is to consider construct validity to be a uniform principle that “subsumes various other aspects of validation” (Cumming and Berwick 1996, 5) and one that “integrates considerations of content, criteria and consequences into a comprehensive framework for empirically testing rational hypotheses about score meaning and utility” (Messick 1995, 742). Although influential in psychometrics, this conflated notion has been contested where empirical data and statistical procedures may not be adequate to account for all aspects involved in measuring language ability (Weideman 2009; Van Dyk 2010; Rambiritch 2012; Du Plessis 2017). Further to this, by considering the social consequences and potential misuse of a test to constitute part of construct validity, as many contemporary language testers do, the theoretical definition of a construct is exceeded. The following more traditional definitions of different kinds of validity remain useful for the purposes of evaluating language assessments (derived from Cumming and Berwick 1996, 1–12; Hughes 2003, 26–35; Weir 2005, 11–15):

- “Criterion-related validity: a combination of concurrent and predictive validity and an aspect that pertains to the correspondence of one set of test results with the results of another test of ability that serves as a criterion;
- Concurrent validity: the ability of the results of one test to correlate with those of another criterion believed to indicate the same ability as that which has been tested;
- Predictive validity: the ability of the results of one test to predict performance in other situations and contexts;
- Construct validity: the theoretical trait or construct of a cognitive and linguistic nature presumably measured, and the alignment thereof with theories on language processing and communicative competence;
- Content validity: the adequate representation of authentic language-related tasks and content in a test.”

(Du Plessis 2017, 73)

The use of assessment data for instructional purposes is traditionally referred to as assessment for learning or formative assessment (Green 2014, 14). In order to inform course content, assessment data have to be interpreted to determine to what extent the students display the

abilities that have been conceptualised theoretically. This affords the opportunity to ensure that sufficient attention is devoted on the teaching side to where it is needed most.

Language learning is dependent on the alignment of context, goals, language learner characteristics and resources (De Graaff and Housen 2011, 728). In language testing terms this would be analogous to exploring the validity of constructs and content and the appropriateness of tasks and item types, while accommodating practical constraints. Determining the relevance of what we do on both the teaching and testing side is thus necessary if we are to be accountable practitioners.

## **METHODOLOGY**

The aim of the current study is primarily to provide a theoretical justification for the design of the language course. Although the investigation may provide an indication of the effectiveness of the module, determining whether instruction has had the desired effect depends on far more than analyzing and interpreting test and assignment marks. Many variables can affect the performance of students, even if the course is well designed and presented by suitably skilled facilitators. The main research question that the study seeks to address is whether the focus of the course falls where it is most needed, in other words whether the right language abilities are being addressed.

Data on the student population are obtained from official student records and a survey questionnaire on literacy habits. This information is supplemented by an analysis of course objectives and curriculum content and the correlation of assessment data. Data are analysed by means of SAS and IteMan 4.3 statistical programming<sup>2</sup> and interpreted on the basis of the course objectives and learning content by applying the language testing notions of construct and content validity, as well as the principle of practicality (Weir 2005; Green 2014).

An important consideration is whether the validation process will lead to positive washback: assessment outcomes should lead to improved teaching practice and skills development – a notion referred to as consequential or systemic validity (Frederiksen and Collins 1989, 27). This is directly related to the impact and power dimension of language assessment (Shohamy 2006). The predicted effects of the assessment are considered to be the impetus that drives the process to design instruments for measuring knowledge and abilities within a “progressive philosophy of language testing” (Fulcher 2015, 126). In terms hereof, the objective of accountability transcends allocating course grades on a reliable and fair basis; it extends to the domain of course refinement for the benefit of students and stakeholders.

## DESCRIPTION OF THE STUDENT POPULATION

The sample population used for the study consisted of a group of 196 education students who intended teaching in the Senior and Further Education and Training (FET) phases. Of these, 24 students failed to meet the course requirements by not completing the necessary assessments. The reasons for this are unknown, and these students were excluded for the purposes of analyzing the assessment data. Of the remaining students, 76 voluntarily participated in a survey on literacy habits. The following table illustrates the diversity of the student survey cohort in respect of first (“home”) languages<sup>3</sup> and National Senior Certificate (NSC) scores for English.

**Table 1:** Pre-course language proficiency levels indicated by NSC results

First language (L1)	Afrikaans	English	Sotho	Tswana	Xhosa	Zulu	Other	Grand Total
<b>Grade 12 English at L1 level</b>	<b>5</b>	<b>3</b>	<b>2</b>	<b>1</b>	<b>5</b>	<b>1</b>	<b>0</b>	<b>17</b>
40–49%	0	0	0	0	1	0	0	1
50–59%	2	1	0	0	2	1	0	6
60–69%	2	2	1	1	1	0	0	7
70–79%	1	0	1	0	0	0	0	2
90–100%	0	0	0	0	1	0	0	1
<b>Grade 12 English at L2 level</b>	<b>36</b>	<b>0</b>	<b>13</b>	<b>1</b>	<b>5</b>	<b>2</b>	<b>1</b>	<b>58</b>
40–49%	0	0	0	0	1	0	0	1
50–59%	1	0	6	1	0	0	0	8
60–69%	11	0	6	0	1	1	1	20
70–79%	17	0	1	0	3	1	0	22
80–89%	7	0	0	0	0	0	0	7
<b>Grade 12 English level not indicated</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>1</b>
60–69%	0	0	0	0	0	1	0	1
<b>Number of students</b>	<b>41</b>	<b>3</b>	<b>15</b>	<b>2</b>	<b>10</b>	<b>4</b>	<b>1</b>	<b>76</b>

We see that the survey cohort represents at least seven different language groups. More than half (54%) are L1 speakers of Afrikaans; only three students (4%) are L1 speakers of English. Three-quarters of the students (76%) studied English at first additional language (FAL) level at school. Of the few students who offered English at L1 level, only three (4%) attained a final mark of 70 per cent or higher. The data illustrate the divergent levels of language proficiency to be addressed in one generic course. Students scoring as low as 40–49 per cent for English at L2 level are required to complete the same module as students achieving more than 70 per cent for English at L1 level. In terms of the principles delineated in the Common European Framework of Reference (CEFR), widely used in the UK, Europe and Asia (Council of Europe 2018), students should attain a certain level of proficiency prior to being admitted to tertiary

study. However, in South Africa the tendency is to admit students to higher education on the basis of their NSC results, and subsequently to require them to complete academic literacy or language courses.

If we were to place students according to their current levels of proficiency (on a similar basis to the CEFR levels), in practical terms this would mean offering English modules on at least three levels (intermediate/low, upper intermediate/moderate and advanced/high proficiency). Unfortunately, insufficient financial and human resources preclude this. Our learning context is restricted to a one-year 16-credit course on NQF level 5, with a maximum of two hours of contact learning per week. No provision is made for successive courses. The fact that one generic course is expected to serve the needs of all students, and that in only 160 notional hours, already weakens the argument for its potential validity and usefulness.

The survey of the literacy habits of students further showed that at least 71 per cent of the respondents regularly took notes in class, but only half summarised course work habitually. A disturbing revelation was that over 67 per cent (around two-thirds) indicated that they did not prepare regularly for classes.

**Table 2:** Learning facilitation habits of students

	Always	Most times	Never	No response	Sometimes	Grand Total
Take notes in class	22	32	3	1	18	76
Write summaries of course work	14	23	7	0	32	76
Prepare before class	3	22	10	0	41	76

The survey also revealed that very few students made a regular habit of undertaking recreational reading not related to academic coursework. Around half indicated that they read books, magazines or newspaper articles only once or twice a month. Close to 20 per cent never read books, online articles or printed newspaper articles.

**Table 3:** Non-academic reading habits

	Every day	A few times a week	Once a week	Once or twice a month	Never	No response	Grand Total
Books	2	17	7	34	15	1	76
Magazines	4	8	17	38	8	1	76
Online articles	6	10	16	29	14	1	76
Printed newspaper articles	3	6	16	34	14	3	76

The literacy profile suggests that students may not have a broad general knowledge and vocabulary to draw on, a matter that could undermine their ability to engage in educational discussions and debates with their own students. It is also dubious whether they would

encourage their students to read extensively, since they themselves appear not to have adopted healthy literacy habits. Not only do educators need to be strong readers themselves, but they also need to be able to lead their students to develop a high level of reading literacy (Kerr and Frese 2017; Murphy, Conway, Murphy and Hall 2014).

## **SUMMARY OF COURSE OBJECTIVES AND OUTLINE**

The information obtained from the literacy survey should be reflected in the course objectives and content. The emphasis of the course under scrutiny is predominantly on the ability to understand texts on a broad range of topics in different formats, and to articulate an appropriately formulated response to information. The specific learning outcomes are listed as follows:

“On completion of the module, you should be able to:

1. understand a range of vocabulary and idiomatic expressions;
2. identify and understand the functions of discourse markers in texts;
3. use your new-found knowledge of grammatical constructions as a vehicle for accomplishing a variety of communicative tasks;
4. express your opinions about a variety of issues fluently, critically and creatively.
5. address different audiences in spoken English with confidence and eloquence.”

(University of the Free State 2015, 3).

The conceptual framework for the course derives from linguistic ideas emanating from the early 1970s on a differentiated communicative competence (Habermas 1970; Hymes 1972; Halliday 1978), and the teaching approach can be described as communicative with text- and task-based components (see Norris 2011; Robinson 2011). Use is made of themed units, and performance tasks are aligned with real-world activities that educators will need to execute in school settings. One hour of contact time each week is devoted to the reading and discussion of a text on a relevant topic. Brief grammatical explanations of forms used in the text are provided and exercises are completed in which these forms are applied and new vocabulary reinforced. A second class each week is designated specifically for the development of oratory skills (learning outcome 5 above). Because students attending the course would be teaching diverse school subjects, topics selected include issues related to culture and identity, as well as health and environmental concerns. The motivation here is that subject matter should represent both the social and natural sciences and be relevant to most students, an important part of the notion of content/context validity (Weir 2005).

## ANALYSIS OF COURSE ASSESSMENT DATA

The analysis of assessment artefacts and data is necessary to determine the extent to which course objectives, instruction and assessment are aligned with students' learning needs.

### Pre- and post-course test results

A multiple-choice test, the Test of English Language Skills (TELS), was developed as a pre- and post-course assessment that would provide a reliable indication of whether students had shown progress in their general language proficiency. For practical reasons the test was designed to be completed within 50 minutes, the normal length of a class. It contained six subtests and 37 test items together contributing 50 marks. Although it was developed as a class test and did not have any pre-piloting, it showed a Cronbach reliability of 0.8. Where a test constitutes one of several types of information on the basis of which inferences of ability are made, an *alpha* value of 0.8 is desirable (see Hogan 2007, 149–150, for a more detailed explanation).

**Table 4:** Reliability of TELS

	Alpha	SEM	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	S-B Random	S-B First-Last	S-B Odd-Even
Pre-test	0.78	2.551	0.590	0.405	0.707	0.742	0.576	0.828
Post-test	0.81	2.572	0.657	0.513	0.733	0.793	0.678	0.846

Because the module was aimed at improving general English proficiency, the test measured abilities based on core objectives 1–3 outlined above. Abilities relating to objectives 4 and 5 were assessed through other assessment artefacts during the course of the year. The summary statistics of the TELS calculated with the Iteman 4.3 statistical software are indicated in Table 5.

**Table 5:** Summary statistics of the TELS pre- and post-test

Score	Items	Admin	Mean	SD	Min Score	Max Score	Mean P	Mean Rpbis
All items	37	Pre-test (N=75)	19.363	5.477	9	33	0.523	0.234
		Post-test (N=165)	21.339	5.849	1	34	0.577	0.282
Text relations	5	Pre-test	2.667	1.506	0	5	0.533	0.302
		Post-test	2.861	1.502	0	5	0.572	0.292
Syntax	7	Pre-test	5.007	1.243	0	7	0.715	0.038
		Post-test	5.224	1.290	0	7	0.746	0.217
Reading graphs	5	Pre-test	2.378	1.292	0	5	0.476	0.197
		Post-test	3.024	1.120	0	5	0.605	0.218
Distinguishing between genres	5	Pre-test	3.578	1.484	0	5	0.716	0.380
		Post-test	3.806	1.530	0	5	0.761	0.413

Score	Items	Admin	Mean	SD	Min Score	Max Score	Mean P	Mean Rpbis
Knowledge of cohesion and coherence	5	Pre-test	2.193	1.655	0	5	0.439	0.324
		Post-test	2.194	1.615	0	5	0.439	0.316
Understanding texts	10	Pre-test	3.541	2.040	0	10	0.354	0.236
		Post-test	4.230	2.137	0	10	0.423	0.273

If we compare the means obtained for the pre- and post-tests, overall students fared somewhat better in the post-test. However, their reading comprehension scores (42% mean) and knowledge of cohesion and coherence (44% mean) remained low. This is also evident in the low facility values (mean P). Disturbingly, in each of the subtests there were students who failed to earn any marks whatsoever (min. score).

In order to determine any improvement in proficiency, a series of statistical procedures was carried out. Because not all students completed all of the assessments, there is missing data that could bias the results of the analysis. For the purposes of mitigating this problem, different types of repeated measure analysis were performed. First, a repeated measures model was fitted with all available assessments (2 assignments, 3 tests, 3 orals and TELS pre- and post-tests) enabling 10 repeated measurements. Hereby an estimate was obtained of the TELS post-test minus TELS pre-test mean difference, together with a confidence interval and p-value. According to this model the mean improvement in proficiency was 4.3 per cent ( $p=0.0051$ ).

**Table 6:** Repeated measures analysis with all available assessments

Least Squares Means									
Effect	Type	Estimate	Standard error	DF	t Value	Pr >  t	Alpha	Lower	Upper
Type	Assignment_1	61.2145	1.4038	177	43.61	<.0001	0.05	58.4443	63.9848
Type	Assignment_2	69.0817	1.2975	182	53.24	<.0001	0.05	66.5216	71.6417
Type	Oral_1	82.5068	0.7510	175	109.86	<.0001	0.05	81.0245	83.9890
Type	Oral_2	74.6264	0.8079	173	92.37	<.0001	0.05	73.0317	76.2210
Type	Oral_3	69.6347	0.6774	174	102.80	<.0001	0.05	68.2977	70.9717
Type	TELS_POST	56.9343	1.1570	174	49.21	<.0001	0.05	54.6507	59.2179
Type	TELS_PRE	52.6057	1.4552	111	36.15	<.0001	0.05	49.7220	55.4893
Type	Test_1	71.5350	1.0619	187	67.37	<.0001	0.05	69.4402	73.6298
Type	Test_2	66.1797	0.8733	188	75.78	<.0001	0.05	64.4570	67.9024
Type	Test_3	67.1654	0.9986	183	67.26	<.0001	0.05	65.1953	69.1356

Differences of Least Squares Means										
Effect	Type	_Type	Estimate	Standard error	DF	t Value	Pr >  t	Alpha	Lower	Upper
Type	TELS_POST	TELS_PRE	4.3287	1.5166	120	2.85	0.0051	0.05	1.3259	7.3315

Hereafter, a repeated measures model with only the TELS pre- and TELS post-tests (two repeated measurements) was fitted. The mean improvement was 4.4 per cent ( $p=0.0064$ ).

**Table 7:** Repeated measures analysis with only the TELS pre- and post-tests

Least Squares Means									
Effect	Type	Estimate	Standard error	DF	t Value	Pr >  t	Alpha	Lower	Upper
Type	TELS_POST	57.1172	1.1877	161	48.09	<.0001	0.05	54.7718	59.4627
Type	TELS_PRE	52.7093	1.4615	99	36.06	<.0001	0.05	49.8093	55.6093

Differences of Least Squares Means										
Effect	Type	_Type	Estimate	Standard	DF	t Value	Pr >  t	Alpha	Lower	Upper
Type	TELS_POST	TELS_PRE	4.4079	1.5839	107	2.78	0.0064	0.05	1.2680	7.5479

Lastly, a procedure similar to that of a paired samples t-test (N=75) was performed, showing a mean improvement of 5.0 per cent (p=0.0069).

**Table 8:** SAS statistical analysis of increased language proficiency

Estimates									
Label	Estimate	Standard error	DF	t Value	Pr >  t	Alpha	Lower	Upper	
Difference TELS_POST-TELS_PRE	5.0133	1.8042	74	2.78	0.0069	0.05	1.4183	8.6084	

The analyses in Tables 6–8 are similar and show a statistically significant but modest improvement in general language proficiency of around 4–5 per cent points. If we examine the constructs assessed, the scores in the tasks that required the understanding of different texts remained low (Table 5). This is where the notions of construct and content validity are particularly informative. If the ability to understand texts is closely related to language proficiency, insufficient attention was devoted to the development of reading skills. Just as the language testing principle of content validity requires sufficient test tasks that involve the reading and processing of information, by analogy the language module would need to include more components that support reading comprehension in order to benefit the students.

### Descriptive statistics of course assessments

If we examine the descriptive statistics of all assessments, we see that the assignments, formative and summative tests, and two of the orals have similar levels of average achievement (around 65–75%). The first oral appears to have measured a different kind of ability to the rest of the oral tasks, producing a much higher mean score.

Table 9 reveals a number of discrepancies with the class mean varying across assessments by as much as 29 per cent. Clearly, the TELS was far more difficult for the students than any of the other assessments. Since the TELS was the only assessment artefact that was scored objectively and analysed statistically, it serves as a reliable benchmark for the remaining

assessments. If assessment artefacts are designed in accordance with the mentioned principles of construct and content validity, there should be a measure of concurrent validity between tests of similar abilities. Table 10 compares the respective subtests and abilities assessed by the TELS with those of the formal tests of general English proficiency.

**Table 9:** Descriptive statistics of all assessment opportunities

	Assign. 1	Assign. 2	Oral 1	Oral 2	Oral 3	Form. Test 1	Form. Test 2	Summ. Test 3	TELS_PRE	TELS_POST	Diff_TELS	Final mark
<b>N</b>	173	180	174	172	166	186	184	178	91	158	75	172
<b>Mean</b>	61.3	69.2	82.6	74.9	69.7	71.7	66.2	67.3	53.0	57.1	5.0	68.0
<b>Std</b>	18.6	17.4	9.9	10.6	8.9	14.5	11.9	13.4	14.6	15.0	15.6	10.3
<b>Min</b>	12.0	24.0	57.5	51.3	44.0	28.0	25.0	30.5	24.0	14.0	-40.0	39.0
<b>Q1</b>	48.0	62.0	75.0	65.0	66.0	62.0	59.0	60.0	44.0	46.0	-2.0	62.5
<b>Median</b>	67.0	73.5	85.0	77.5	70.0	72.0	67.0	69.0	52.0	57.0	4.0	69.0
<b>Q3</b>	75.0	81.0	90.0	82.5	76.0	82.0	74.5	77.0	62.0	68.0	14.0	75.5
<b>Max</b>	90.0	96.0	100.0	97.5	90.0	98.0	93.0	94.0	88.0	88.0	52.0	89.0

Common to each of the four tests are the ability to understand texts, knowledge of and ability to use discourse markers (text relations), and knowledge of grammar. In order to glean more useful information to support the validity of constructs on the teaching and assessment side, scores per subtest need to be computed separately. Since only the total test mark per student was recorded for the three formal course tests, it is impossible to correlate scores for any of the subtests with those of the TELS. The ability to understand texts was identified as an area of weakness in the TELS, but we cannot make the same claim in the case of the remaining tests without knowing the scores per subsection. It would furthermore be desirable to compute the marks allocated per item for each student so as to calculate the reliability coefficients per assessment opportunity. Unfortunately, this is usually not practically feasible, hence the need for a baseline assessment such as the TELS.

**Table 10:** General language proficiency subsections used in course tests

TELS	Items	Formative Test 1	Items	Formative Test 2	Items	Summative Test	Items
Text relations	5	Word formation	5	Understanding texts	12	Understanding texts	16
Syntax	7	Understanding texts	9	Paragraph writing task	1	Paragraph writing task	1
Reading graphs	5	Paragraph writing task	1	Grammar and text relations	9	Grammar and text relations	11
Distinguishing between genres	5	Grammar and text relations	10				
Knowledge of cohesion and coherence	5	Text editing	5				
Understanding texts	10						
<b>Total items</b>	<b>37</b>	<b>Total items</b>	<b>30</b>	<b>Total items</b>	<b>22</b>	<b>Total items</b>	<b>28</b>

In comparison to the low averages for the TELS, the notably high averages for oral work suggest that there may have been a lack of conceptual clarity of constructs. Here too data were not available to determine inter-rater correlations and scoring consistency. However, all academic staff members involved in the oral assessments agreed that students showed confidence in speaking English and that most had performed well in this curriculum component. In terms of generalizability of ability to non-testing domains, the oral tasks demonstrated content and construct validity as assessment artefacts of general communicative competence. The written tests, on the other hand, assessed cognitive academic language proficiency (CALP) and not basic interpersonal communication skills (BICS) (Cummins and Davison 2007). This would explain the discrepancy in results between oral and written assessments, and points to a need for greater alignment of standards and levels across all course components.

### **Correlation of all course assessment data**

Correlation of all assessment data enables the identification of closely related variables. For the purposes of correlating data, marks have been re-worked as a percentage. This enables comparisons but does not change the nature of the correlations.

The correlation of the TELS pre- and post-tests with the final composite course mark is relatively high (0.41 and 0.67 respectively), indicating concurrent validity between the assessment instruments: the shorter TELS serves as a reliable measure of general language proficiency in a similar manner to the composite final mark attained. We also see that, with the exception of the first two oral assessments, all formative and summative assessments are quite strongly correlated to the TELS score and final composite course mark. Cohen (1988, 79–81) considers a correlation of .30 to .49 to indicate medium strength, while a correlation of .50 or more between variables would be indicative of a strong correlation. The strength of the relationship is important in terms of the reliability and usefulness of the assessment results. This also matters in respect of the notion of consequential validity. Most of the course assessments correlate quite strongly with the final course outcome, which suggests that the assessment was reliable, valid and fair towards students.

Viewed in isolation, the correlation of assessment data provides insufficient information for the theoretical justification of the course design. When we interpret the course results on the basis of the objectives and learning constructs, it becomes clear that not all of the core objectives were met, especially in respect of critical literacies.

**Table 11:** Statistical correlation of course assessment data

Pearson Correlation Coefficients Prob >  r  under H0: Rho=0 Number of Observations												
	Assign_1	Assign_2	Oral_1	Oral_2	Oral_3	TELS_POST	TELS_PRE	Test_1	Test_2	Test_3	Diff_TELS	Final_mark
<b>Assign_1</b>	1.00	0.16489	0.15214	0.15991	0.13686	0.28139	0.02885	0.29762	0.22719	0.23623	0.36213	0.48607
		0.0332	0.0518	0.0421	0.0895	0.0006	0.7932	<.0001	0.0030	0.0023	0.0016	<.0001
	173	167	164	162	155	146	85	170	169	165	73	162
<b>Assign_2</b>	0.16489	1.00000	0.02475	0.01190	0.15637	0.16582	0.16235	0.18413	0.32579	0.30141	-0.12315	0.56424
	0.0332		0.7509	0.8795	0.0456	0.0398	0.1401	0.0147	<.0001	<.0001	0.2993	<.0001
	167	180	167	165	164	154	84	175	178	174	73	169
<b>Oral_1</b>	0.15214	0.02475	1.00000	0.60112	0.23608	0.24320	0.04107	0.20334	0.23693	0.20174	0.09065	0.31045
	0.0518	0.7509		<.0001	0.0028	0.0029	0.7107	0.0078	0.0019	0.0091	0.4522	<.0001
	164	167	174	172	158	148	84	170	170	166	71	162
<b>Oral_2</b>	0.15991	0.01190	0.60112	1.00000	0.13403	0.23865	0.08965	0.19517	0.23446	0.28847	0.18351	0.31127
	0.0421	0.8795	<.0001		0.0932	0.0036	0.4174	0.0112	0.0022	0.0002	0.1255	<.0001
	162	165	172	172	158	147	84	168	168	164	71	160
<b>Oral_3</b>	0.13686	0.15637	0.23608	0.13403	1.00000	0.39189	0.28689	0.41858	0.41331	0.36650	0.10811	0.48354
	0.0895	0.0456	0.0028	0.0932		<.0001	0.0109	<.0001	<.0001	<.0001	0.3838	<.0001
	155	164	158	158	166	144	78	162	164	162	67	157
<b>TELS_PO</b>	0.28139	0.16582	0.24320	0.23865	0.39189	1.00000	0.45530	0.52946	0.47432	0.57095	0.57721	0.67284
	0.0006	0.0398	0.0029	0.0036	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
	146	154	148	147	144	158	75	153	156	152	75	147
<b>TELS_PR</b>	0.02885	0.16235	0.04107	0.08965	0.28689	0.45530	1.00000	0.32952	0.44437	0.43264	-0.46424	0.41120
	0.7932	0.1401	0.7107	0.4174	0.0109	<.0001		0.0017	<.0001	<.0001	<.0001	0.0001
	85	84	84	84	78	75	91	88	86	84	75	83
<b>Test_1</b>	0.29762	0.18413	0.20334	0.19517	0.41858	0.52946	0.32952	1.00000	0.57318	0.56363	0.19221	0.59122
	<.0001	0.0147	0.0078	0.0112	<.0001	<.0001	0.0017		<.0001	<.0001	0.1009	<.0001
	170	175	170	168	162	153	88	186	178	174	74	172
<b>Test_2</b>	0.22719	0.32579	0.23693	0.23446	0.41331	0.47432	0.44437	0.57318	1.00000	0.56084	-0.06496	0.60297
	0.0030	<.0001	0.0019	0.0022	<.0001	<.0001	<.0001	<.0001		<.0001	0.5824	<.0001
	169	178	170	168	164	156	86	178	184	175	74	171
<b>Test_3</b>	0.23623	0.30141	0.20174	0.28847	0.36650	0.57095	0.43264	0.56363	0.56084	1.00000	0.09462	0.69566
	0.0023	<.0001	0.0091	0.0002	<.0001	<.0001	<.0001	<.0001	<.0001		0.4259	<.0001
	165	174	166	164	162	152	84	174	175	178	73	172
<b>Diff_TELS</b>	0.36213	-0.12315	0.09065	0.18351	0.10811	0.57721	-0.46424	0.19221	-0.06496	0.09462	1.00000	0.28837
	0.0016	0.2993	0.4522	0.1255	0.3838	<.0001	<.0001	0.1009	0.5824	0.4259		0.0140
	73	73	71	71	67	75	75	74	74	73	75	72
<b>Final mark</b>	0.48607	0.56424	0.31045	0.31127	0.48354	0.67284	0.41120	0.59122	0.60297	0.69566	0.28837	1.00000
	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0001	<.0001	<.0001	<.0001	0.0140	
	162	169	162	160	157	147	83	172	171	172	72	172

## Results of a pilot study on critical literacy as a construct in language learning

The inadequate attention given to general reading proficiency as a learning outcome has already been pointed out. Apart from this shortcoming, the analysis of curriculum content and assessment data showed that the following critical outcomes were neglected:

- “identify and solve problems and make decisions using critical and creative thinking;
- collect, analyze, organise and critically evaluate information;
- demonstrate an understanding of the world as a set of related systems by recognising that problem-solving contexts do not exist in isolation.”

(University of the Free State 2015, 4).

Although the tests included some inferential reading, the content analysis showed that this was insufficient to develop the kind of critical thinking necessary for academic language proficiency (Grosser and Nel 2013), an area in which South African students tend to perform poorly (Rademeyer 2007). To investigate this aspect further, a small pilot study was undertaken of responses to a reading comprehension assignment identified as suitable for examining the notion of critical language awareness (CLA). Convenience sampling was used as this task had been completed by a small group of 30 students as a further assessment opportunity.

In order to contextualise CLA, the definition of Language Awareness (LA) should first be explored. Kiely (2009, 331, after Svalberg 2007), identifies two salient characteristics of LA; the first of which refers to a “broad-based movement, embracing both research and pedagogic action”. Its goal is to harness emerging understanding in the applied linguistics and educational fields and to identify patterns and practices that may assist in the learning process. Secondly, LA has morphed over the years to encompass new and emerging investigative branches in the field. The original focus on grammar and language form in language learning has been extended to include cognitive and social dimensions of language learning and use, bearing testimony to the “holistic nature” of the LA project (Svalberg 2007, 287). LA thus highlights the synergy between various topics of critical pedagogy, power and language use in society, while various degrees of understanding (cf. the differentiation created between *noticing* and *understanding*) have become more pertinent (Svalberg 2007, 287–290). This is relevant with regards to the development of formal aspects of academic literacy. Many of the features that add value to text do not derive from the application of specific rules, as Kiely (2009) notes. They relate to a sense of appropriateness in the specific textual context, and they are features that students have to notice and then understand. Both Janks (2000, 2001, 2010) and Cummins (2001) advocate the

development of mainstream academic language and the establishing of CLA for deep understanding, albeit from different angles. Janks' (2010) CLA approach emanates from a focus on ideological and textual analysis, whereas Cummins (2001) focuses on language and meaning in his Academic Expertise Framework.

We support Shamim's (2011, 30) definition of LA, as "a mental and internal capacity which the learner gradually develops by giving motivated and conscious attention to language in use to discover its patterns". CLA derives from LA and increases learners' comprehension of language functions. The major function of LA is that learners "find out language" for themselves (Shamim 2011, 30). In the process a spirit of investigation is fostered in students. This is relevant to our multicultural contexts that require specific communicative ability and knowledge: "... an awareness of how social and cultural categories intertwine and evolve both inside and outside the classroom is believed to be central to such knowledge" (Orgeret 2012, 201).

Blommaert's (2005, 65) question, "Why do we investigate this now?", is especially relevant to the social context of our own research. The preliminary micro-study data retrieved from the corpus of written responses strongly suggests a xenophobic tendency amongst the students of our own institution. These findings correlate with studies at other universities (cf. Singh, 2013, 88 and Moge kwu, 2005, 10) where one would also expect there to be a high level of open-mindedness, especially amongst students being trained to be educators. Responses revealed students' attitudes towards non-nationals and xenophobia. Answers were sorted according to three attitudinal categories; namely a negative category (indicating negativity towards non-nationals), neutral and positive categories. Responses that indicated prejudice towards non-nationals were categorized as negative. Incorrect responses were regarded as neutral, as well as responses that did not use emotional or prejudicial language or ideas. Ideas that favoured the presence of non-nationals, or that conveyed sadness or grief because of the violent nature of xenophobia, were seen as indicative of an ability to identify or empathize with the plight of non-nationals; therefore, justifying the categorizing of these responses as more positive attitudinal responses.

**Table 12:** Summary of results of CLA pilot study

<b>Positive</b>	<b>Negative</b>	<b>Neutral</b>	<b>Incorrect</b>
23.3% of student responses	46.6% of student responses	23.3% of student responses	6.6%
7 responses	14 responses	7 responses	2 responses

In brief, the dominant notion amongst students was that African migrants were taking advantage of South African hospitality: they were the cause of the high crime rate in South Africa,

economic hardships, unemployment, and a plethora of other social ills – notions also held by the broader South African public, as explicated by Neocosmos (2008, 590). Discourse analysts such as Fairclough believe critical awareness of the role of discourse is required for personal success and social change in society, and that it is the role of language education to promote such awareness. According to Fairclough (1995, 222), CLA is a prerequisite for effective citizenship and a democratic way of life. Critical Discourse Analysis (CDA) is thus one approach that can be utilised in language education to develop critical reading literacies. In fact, a study by Shamim (2011, 28–35) indicated a 90 per cent increase in students' CLA after they were introduced to CDA techniques. This is therefore a notion that needs to be incorporated in any revised language curriculum, especially where close reading abilities need development. Here too, the analysis of course assessment artefacts was useful to evaluate the amount of attention devoted to critical literacies in the module.

## **CONCLUDING THOUGHTS**

This article has attempted to highlight the importance of employing different kinds of data to inform course content and instructional practice with a view to ensuring that learning objectives are adequately met, including the development of critical literacies needed for responsible citizenry and social change. The results of the pilot reading assignment enhance the existing body of research on critical language awareness and reveal “the ideological bases of discourse as they circulate both in everyday life and within significant texts” (Flowerdew 2013, 192), as produced by students in a South African Higher Educational setting. In this respect, the term “informative assessment” illustrates a double meaning at play: not only can the correlation, analysis and interpretation of assessment data provide useful feedback to students on their course progress and facilitate their learning, but it can also be used to shape responsive and accountable pedagogies.

Systemic validity in language teaching is reliant on the analysis and interpretation of data pertaining to course objectives, content and learner performance. Through the triangulation of these kinds of data, it is possible to find theoretical justification for course design and instructional practice. Multiple sources of evidence are necessary, both on the instructional and assessment side, to support the theoretical validity of a course. Further to this, by examining the concurrent validity of course assessments, the consistency of measurement can be monitored, an essential part of the quest to validate what language teachers do and to help them become more reflective practitioners.

Attention to curriculum design and instructional practice is of paramount importance in any teaching context, but particularly so when resources are scarce and have to be used to

eradicate negative legacies of previous political and educational dispensations. Notwithstanding the serious practical constraints applicable to the language module under review, the analysis and correlation of data indicate that the abilities to negotiate discourse, understand texts and develop critical literacies in particular need foregrounding. The pilot study on critical language awareness points to a need for educators and students to display a greater sense of self-consciousness with regards to perceptions of non-nationals, not only in terms of course materials selected, but also in terms of fostering a greater sense of community within the classroom. Overt and covert xenophobic sentiments need to be addressed not only on a micro level in school classrooms, but also on a macro level in higher education institutions across the country. The theme of inequality (and inclusivity) in a broader sense remains a common South African thread that is firmly embedded within our social fibre, and paradoxically includes the “Other” (Said 1978). Following Freirean principles, the language course rationale should become a catalyst for ways of living beyond the English second language classroom.

The study illustrates the necessity of holistic evaluation and triangulation of curriculum and course content, student needs and course performance data. Without this, the actual learning needs of students undergoing training at tertiary institutions may not be addressed, and accountability on the part of those responsible for their education could be said to be lacking, a situation which the country can ill afford.

## NOTES

1. In essence, standardised testing refers to the administration of a test, which has usually been piloted and refined, under the same conditions to different groups of candidates often at different points in time, as well as clearly defined and fixed methods of scoring (Fulcher 2010, 5).
2. The assistance of Prof. Schall and the Statistical Consultation Unit of the University of the Free State with the correlation and interpretation of data is acknowledged.
3. We follow the English writing convention for the names of the Bantu languages.

## REFERENCES

- Blommaert, J. 2005. *Discourse: A critical introduction (Key topics in Sociolinguistics)*. Cambridge: Cambridge University Press.
- CDE see Centre for Development and Enterprise.
- Centre for Development and Enterprise. 2015. Teachers in South Africa: Supply and demand 2013–2025. Executive summary of a report. <http://www.cde.org.za/wp-content/uploads/2015/03/Final-Revised-ES-TeacherSupplyandDemand2025.pdf> (Accessed 14 June 2018).
- Cohen, J. W. 1988. *Statistical power analysis for the behavioural sciences*. 2<sup>nd</sup> Edition. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Council of Europe. 2018. Common European Framework of Reference for Languages (CEFR): Learning, teaching, assessment. Cambridge: Cambridge University Press. [http://www.coe.int/t/dg4/linguistic/Cadre1\\_en.asp](http://www.coe.int/t/dg4/linguistic/Cadre1_en.asp) (Accessed 14 June 2018).

- Cumming, A. and R. Berwick. (Ed.). 1996. *Validation in language testing*. Clevedon, UK: Multilingual Matters Ltd.
- Cummins, J. 2001. *Negotiating identities: Education for empowerment in a diverse society*. 2<sup>nd</sup> Edition. Los Angeles, CA: CAFE.
- Cummins, J. and C. Davison (Ed.). 2007. *International handbook of English language teaching*. New York: Springer Science+Business Media, LLC.
- De Graaf, R. and A. Housen. 2011. Investigating the effects and effectiveness of L2 instruction. In *The handbook of language teaching*, ed. M. H. Long and C. J. Doughty, 726–755. Chichester, West Sussex: Wiley-Blackwell.
- Du Plessis, C. 2012. The design, refinement and reception of a test of academic literacy for postgraduate students. MA dissertation, University of the Free State, South Africa.
- Du Plessis, C. 2017. Developing a theoretical rationale for the attainment of greater equivalence of standard in the Grade 12 Home Language exit-level examinations. PhD thesis, University of the Free State, South Africa.
- Ellis, R. 2008. *The study of second language acquisition*. 2<sup>nd</sup> Edition. Oxford: Oxford University Press.
- Fairclough, N. 1995. *Critical discourse analysis: The critical study of language*. London: Longman.
- Flowerdew, J. 2013. *Discourse in English language education*. London: Routledge.
- Frederiksen, J. R. and A. Collins. 1989. A systems approach to educational testing. *Educational Researcher* 18(9): 27–32.
- Fulcher, G. 2010. *Practical language testing*. London: Hodder Education.
- Fulcher, G. 2015. *Re-examining language testing: A philosophical and social inquiry*. Abingdon, Oxon: Routledge.
- Green, A. 2014. *Exploring language assessment and testing*. New York: Routledge.
- Grosser, M. M. and M. Nel. 2013. The relationship between the critical thinking skills and the academic language proficiency of prospective teachers. *South African Journal of Education* 33(2): 1–17.
- Habermas, J. 1970. Toward a theory of communicative competence. In *Recent sociology 2*, ed. H. P. Dreitzel, 115–148. London: Collier-Macmillan.
- Halliday, M. A. K. 1978. *Language as social semiotic: The social interpretation of language and meaning*. London: Edward Arnold.
- Hogan, T. P. 2007. *Psychological testing: A practical introduction*. 2<sup>nd</sup> Edition. Hoboken, New Jersey: John Wiley and Sons.
- Hughes, A. 2003. *Testing for language teachers*. 2<sup>nd</sup> Edition. Cambridge: Cambridge University Press.
- Hymes, D. 1972. On communicative competence. In *Sociolinguistics: Selected readings*, ed. J. B. Pride and J. Holmes, 269–293. Harmondsworth: Penguin.
- Janks, H. 2000. Domination, access, diversity and design: A synthesis for critical literacy education. *Educational Review* 52(2): 175–186.
- Janks, H. 2001. Identity and conflict in the critical literacy classroom. In *Negotiating critical literacies in classrooms*, ed. B. Comber and A. Simpson, 137–150. Mahwah, NJ: Erlbaum.
- Janks, H. 2010. *Literacy and power*. New York: Routledge.
- Kerr, M. M. and K. M. Frese. 2017. Reading to learn or learning to read? Engaging college students in course readings. *College teaching* 65(1): 28–31. <https://www.tandfonline.com/doi/abs/10.1080/87567555.2016.1222577> (Accessed 14 June 2018).
- Kiely, R. 2009. Observing, noticing, and understanding: Two case studies in language awareness in the development of academic literacy. *Language awareness* 18: 3–4, 329–344. <https://dx.doi.org/10.1080/09658410903197298> (Accessed 14 June 2018).
- Messick, S. 1995. Validity of psychological assessment: Validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American psychologist* 50:

741–749.

- Mogekwu, M. 2005. African Union: Xenophobia as poor intercultural communication. *Ecquid Novi* 26(1): 5–20.
- Murphy, B., P. F. Conway, R. Murphy and K. Hall. 2014. The emergence of reading literacy in post-primary teacher education: From the background to the foreground. *European Journal of Teacher Education* 37(3): 331–347. <https://www.tandfonline.com/doi/abs/10.1080/02619768.2013.870995> (Accessed 14 June 2018).
- Neocosmos, M. 2008. The politics of fear and the fear of politics: Reflections on Xenophobic violence in South Africa. *Journal of Asian and African studies* 43(6): 586–594.
- Nkosi, B. 2015. Most of our new teachers can't. *Mail & Guardian*, 17 April 2015. <http://mg.co.za/article/2015-04-17-most-of-our-new-teachers-cant> (Accessed 20 October 2015).
- Norris, J. M. and L. Ortega. 2000. Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning* 50: 417–528.
- Norris, J. M. 2011. Task-based teaching and testing. In *The handbook of language teaching*, ed. M. H. Long and C. J. Doughty, 578–594. Chichester, West Sussex: Wiley-Blackwell.
- Orgeret, K. S. 2012. Intercultural educational practices: Opening paths for dialogue. *Intercultural Communication Studies* (XXI): 189–204. <http://hdl.handle.net/10642/1379> (Accessed 14 June 2018).
- Rademeyer A. 2007. SA vaar swakste in toets: Leerlinge se leesvaardighede nie op peil. [SA performs the worst in test: Learners' reading abilities not up to standard – own translation]. *Beeld* 30 November.
- Rambiritch, A. 2012. Challenging Messick: Proposing a theoretical framework for understanding fundamental concepts in language testing. *Journal for Language Teaching* 46(2): 108–121.
- Robinson, P. 2011. Syllabus design. In *The handbook of language teaching*, ed. M. H. Long and C. J. Doughty, 294–310. Chichester, West Sussex: Wiley-Blackwell.
- Said, E. 1978. *Orientalism*. New York: Random House.
- Shamim, A. 2011. Critical language awareness in pedagogic context. *Language Teaching* 4(4): 28–35. [www.ccsenet.org/elt](http://www.ccsenet.org/elt) (Accessed 14 June 2018).
- Shohamy, E. 2006. *Language policy: Hidden agendas and new approaches*. New York: Routledge.
- Singh, R. J. 2013. Examining xenophobic practices amongst university students – a case study from Limpopo province. *Alternation (Special Edition)* 7: 88–108.
- Svalberg, A. M. L. 2007. Language awareness and language learning. *English Language Teaching Journal* 40: 287–308.
- University of the Free State. 2015. *English for professional purposes (ENGS1504) course guide*. Bloemfontein: University of the Free State.
- Van Dyk, T. J. 2010. Konstitutiewe voorwaardes vir die ontwerp van 'n toets van akademiese geletterdheid. [Constitutive conditions for the design of a test of academic literacy]. PhD thesis, University of the Free State, South Africa.
- Weideman, A. 2009. Constitutive and regulative conditions for the assessment of academic literacy. *Southern African Linguistics and Applied Language Studies* 27(3): 235–251.
- Weideman, A. 2017a. Does responsibility encompass ethicality and accountability in language assessment? *Language & Communication* 57: 1–9. <http://dx.doi.org/10.1016/j.langcom.2016.12.004> (Accessed 3 August 2017).
- Weideman, A. 2017b. *Responsible design in applied linguistics: Theory and practice*. Cham, Switzerland: Springer.
- Weir, C. J. 2005. *Language testing and validation: An evidence-based approach*. New York: Palgrave Macmillan.