


# Time limits and English proficiency tests: Predicting academic performance

**Author:**Ingrid Opperman<sup>1</sup> **Affiliation:**

<sup>1</sup>Department of Student Development and Support, Higher Education Development and Support, Tshwane University of Technology, Pretoria, South Africa

**Corresponding author:**

Ingrid Opperman,  
ingrid258@gmail.com

**Dates:**

Received: 30 Oct. 2019

Accepted: 02 June 2020

Published: 25 June 2020

**How to cite this article:**

Opperman, I. (2020). Time limits and English proficiency tests: Predicting academic performance. *African Journal of Psychological Assessment*, 2(0), a20. <https://doi.org/10.4102/ajopa.v2i0.20>

**Copyright:**

© 2020. The Authors.  
Licensee: AOSIS. This work  
is licensed under the  
Creative Commons  
Attribution License.

English is the primary language of instruction in South African higher education, but entering students of first year are often not sufficiently proficient. Therefore, a need is evident for proficiency testing to guide intervention initiatives. International proficiency tests are lengthy and expensive, but Cloze procedure and vocabulary tests have been used as effective alternatives. However, time limits may affect observed reliability and predictive validity in the context of higher education. The present research assessed a cohort of first-year tourism management students using versions of the English Literacy Skills Assessment (ELSA) Cloze procedure and Vocabulary in Context tests under three time-limit conditions: normal, double and no time limits. Students in double and no time-limit conditions performed significantly better than the normal time-limit group. Group scores were correlated with, and significant predictors of, academic subject first-test scores. Better performance and more accurate prediction under extended time limits may be related to students attempting more questions. As the ELSA Vocabulary in Context was the better predictor in this research, the importance of non-technical vocabulary, as opposed to semantic and contextual understandings in Cloze procedure, is highlighted. Therefore, screening the English proficiency levels of students admitted to higher education institutions may be useful to flag likelihood of success and guide interventions.

**Keywords:** higher education; English proficiency; Cloze procedure; vocabulary; time limits.

## Introduction

English has become the dominant language of business, public life and higher education (Benzie, 2010; Casale & Posel, 2011; Coleman, 2006; Nunan, 2003). Therefore, formal acquisition of English language skills has become essential for success in both higher education and business contexts to enhance economic opportunities in a multinational and international economy (Bedenlier & Zawacki-Richter, 2015; Prinsloo & Heugh, 2013). Higher education serves an essential role in enhancing the future career prospects in a competitive social and economic framework, making success integral for many young people (Coleman, 2006; Cross & Carpentier, 2009; Prinsloo & Heugh, 2013). Although higher academic success has become essential for entry into the 21st century economy (Jackson, 2015), academic English language proficiency remains a challenge for the majority of South African students in a linguistically diverse society (Andrade, 2006; Cross & Carpentier, 2009; Murray, 2010; Trenkic & Warmington, 2018).

Academic English proficiency in higher education encompasses formal and functional control of the properties of English language, including vocabulary, grammar and contextual understanding (Bridgeman, McBride, & Monaghan, 2004; Masrai & Milton, 2018; Murray, 2010). Limited English proficiency on entry may lead to academic vulnerability, characterised by unsuccessful adaptation to higher education demands, which could be detrimental to academic literacy, problem-solving techniques, constructive engagement in learning processes (Murray, 2010; Taylor & Von Fintel, 2016) and communications (Benzie, 2010; Murray, 2010; Trenkic & Warmington, 2018; Webb, 2002). Concomitantly, lack of capability in basic interpersonal communication skills (BISC; expression of conversational fluency) alongside cognitive academic language proficiency (CALP; decontextualised language proficiency) may synergistically impact the expression of general English language proficiency in multiple contexts (Barton, Wisessuwan, & Tubsree, 2018; Cummins, 2000). This disadvantage is displayed where decontextualised language learning experiences in everyday learning and communications, linked to BISC, impact the learning of academic concepts, and thereby result in less than optimal CALP (Abriam-Yago, Yoder, & Kataoka-Yahiro, 1999; Tomasello, 2014). Thus, students lacking English language skills sufficient for the tertiary academic environment are placed at a disadvantage, even if basic literacy skills are sufficient.

**Read online:**

Scan this QR  
code with your  
smart phone or  
mobile device  
to read online.

Apart from basic literacy, the context of higher education often requires content-specific skills (linked to CALP; Cummins, 2000), which are reliant on technical vocabulary beside general contextual identification and understanding (Dalton-Puffer, 2011; Fenton-Smith, Humphreys, & Walkinshaw, 2018; Millin & Millin, 2018). Global research has implied that basic skills are a necessary component for developing technical/academic language (Birrell, 2006; Coleman, 2006). Consequentially, students lacking English proficiency skills, or exhibiting competency gaps, may be at an academic disadvantage on entering English language institutions.

Internationally, English proficiency tests are frequently conducted pre-admission for selection purposes. Although these tests could be utilised for admitting students in first year, they are often time-consuming, expensive and focused on overall proficiency rather than critical basic skills more relevant to post-admissions phase (Arrigoni & Clark, 2015; Feast, 2002; Goto, Maki, & Kasai, 2010; Murray, 2010). These traditional gate-keeping tests include the International English Language Testing System (IETLS) and the Test of English as a Foreign Language (TOEFL). The viability and financial feasibility of utilising these assessments post-admissions to identify competency gaps is insufficient. Post-admissions, other options, including the Diagnostic English Language Test and Diagnostic English Language Needs Assessment, have been used globally for screening and diagnosis with good predictive and diagnostic validity (Doe, 2014; Read, 2008). Similar to pre-admission tests, the foci include vocabulary, speed-reading, listening and interpretation of texts. In both cases, complex, rather than base skills are inherent to the tests. Thus, other research has indicated that briefer, basic ability tests, including Cloze procedure protocols and vocabulary assessments, are time- and cost-effective whilst retaining sufficient psychometric properties (Goto et al., 2010; Sun & Henrichsen, 2010).

Cloze procedure protocols require the reader to insert missing words or phrases, illustrating semantic and contextual understanding linked to reading comprehension and writing skills (Gellert & Elbro, 2013; Trace, Brown, Janssen, & Kozhevnikova, 2017). Such skills are considered essential in higher education and significantly vulnerable for second-language English speakers, perhaps because of inability to decode new information and translate key words within specific contexts (Escamilla, 2009; Huettig, 2015; Staub, Grant, Astheimer, & Cohen, 2015). Decoding, recognition and translation to English (in the case of non-native speakers) have been closely related to Cloze procedure protocol performance in children and adults (Gellert & Elbro, 2013; Keenan, Betjemann, & Olson, 2008). These findings suggest that background and fundamental learning could play a role in developing essential skills which are transferable to higher education English language requirements. Similarly, vocabulary acquisition has been linked to success in the context of higher education.

Acquired vocabulary has often been used as a proxy for general proficiency, demonstrating predictive power (Masrai & Milton, 2018; Trenkic & Warmington, 2018). Non-technical vocabulary levels have been further linked to academic writing, reading comprehension and general academic performance (Harrington & Roche, 2014; Qian, 2002; Schmitt, Jiang, & Grabe, 2011; Snow, Lawrence, & White, 2009; Trenkic & Warmington, 2018). These findings are supportive of the inclusion of vocabulary components in traditional gate-keeping tests, lending support for the use of these tests as a proxy for proficiency even post-admissions in first-year students. In both cases, the feasibility of reduction in time and cost is a significant benefit.

Although research has demonstrated that both Cloze procedure protocols and contextually based vocabulary tests may be used as proxies to understand English proficiency, these assessments are often conducted under time constraints, potentially confounding content performance with response time (e.g. Goto et al., 2010; Harrington & Roche, 2014; Masrai & Milton, 2018). Administration under time-constrained conditions remains a common practice for a variety of reasons but may result in decreased validity and reliability values (Van der Linden, 2011). Concomitantly, the test may then lack accuracy for its stated purpose, which is problematic for both selections and post-admission competency identification contexts. Therefore, a balance between internal consistency, predictive validity, length of assessment and other administration factors is required to enhance identification of the status of English language skills. The question then arises as to whether a sufficient balance of time-effectiveness, practicality and predictive validity is present when time constraints are implemented.

Researchers have reported improvements in performance on various English language tests with additional time allocations (Bridgeman et al., 2004; Powers & Fowles, 1997), suggesting a focus on performance in complex understandings may be more important for academic outcomes than time-constrained responses (Daly & Stahmann, 1968; Harrington & Roche, 2014; MacIntyre & Gardner, 1994). The removal of time constraints may also mitigate other factors associated with poorer performance, including inadequate test-taking strategies, test anxiety and familiarity with testing contexts (Anderson, 1991; Fairbairn, 2007; Solano-Flores, 2008). Similar findings are present in the context of higher education, for which increased predictive validity, reliability and construct validity of Cloze procedure protocols and vocabulary tests have been reported when time constraints are removed (Hajebi, Taheri, & Allami, 2018; Snow et al., 2009; Trace et al., 2017).

Researchers have hypothesised that changes in performance under different time constraints may be linked to the number of items attempted, changes to item structures or content functions operating differently (Luke & Christianson, 2016; Talento-Miller, Guo, & Han, 2013; Van der Linden, 2011). Other research has suggested that increased time may allow for better translation and internal reconstructions of

semantics and syntax, although this may only be true for lengthy fragments in Cloze procedure protocols or when a wide range of possible responses is presented (Hajebi et al., 2018; Staub et al., 2015). Although this research has considered Cloze procedure protocols, vocabulary and other English proficiency tests without time constraints, limited published work (e.g. Goto, Maki, & Kasai, 2010) has considered different predictive validity of short assessments under various time constraints.

The present study assessed the relative influence of time limits on two English language proficiency tests, that is, a Cloze procedure protocol and contextual vocabulary assessment, to understand differences in the predictive validity under each time limit in determining first-test academic outcomes. The importance of this study lies in differentiation between English proficiency itself and the impact of time constraints on the expression of that proficiency in predicting academic outcomes. Thus, the study intends to contribute through further understanding English proficiency testing in terms of the potentially detrimental impact of time limitations on test outcomes. These findings are potentially useful in enhancing mass language post-admission screening to improve skills-targeted interventions which are time-efficient and effective.

## Methods

### Participants

Participants comprised commencing first-year students ( $n = 81$ ) enrolled in an institute for a tourism management national diploma course with common first-year academic subjects and admission requirements. The restriction for course enrolment was intended to indirectly standardise minimum English language entry criteria. The majority of enrolled first-year students at the institute were aged between 18 and 20 years, with a vast majority being of black ethnicity equally split between males and females.

### Research design

The present research made use of a cross-sectional, quasi-experimental design to assess the impact of different time limits on performance of both Cloze procedure protocol and contextual vocabulary assessment.

### Instruments

Kaleidoprax (2014) developed English Literacy Skills Assessment (ELSA) as two modified tests for the institute conducting the study: the Cloze procedure and the Vocabulary in Context tests. At present, no psychometric properties have been made available for the tests (Kaleidoprax, 2014). The Cloze procedure test requires the insertion of missing words within the context of a sentence. Cloze procedure comprises 20 questions, each with four possible responses, of which one is correct ( $\text{max} = 20$ ). The Vocabulary in Context test identifies words in the context of a full sentence to require extrapolation of meaning of definitions, synonyms, antonyms and usage.

Vocabulary in Context comprises 30 questions, each with four possible responses, of which one is correct ( $\text{max} = 30$ ). No penalty scoring is implemented for either test. In this study, academic performance was assessed using percentages for the first-test marks for first-year subjects of national diploma courses in the department of tourism management ( $\text{min} = 1\%$ ,  $\text{max} = 100\%$ ). All marks obtained were above 0%.

### Procedure

Data on the ELSA were generated as part of administration of a battery which took place after English language portion. The battery was solicited by the academic departments of the institute as part of a post-admissions first-year student assessment. Academic departments granted permission to modify the English language portion for research purposes, and all participants gave informed consent. No data were used for exclusionary, probationary or placement purposes.

The full sample ( $n = 81$ ) was broken down into three groups: Normal time limit ( $n = 44$ ), double time limit ( $n = 23$ ) and no time limit ( $n = 15$ ). Separate test sessions took place for each group. Participants had freedom to join the group of their choice. Participation in the experimental group was voluntary, and verbal informed consent was obtained with written signatory. Because of the voluntary nature of participation, a convenience sample was produced. Resultantly, control for Grade-12 English performance and the size of groups were not possible. Voluntariness of participation, however, was essential because of the testing (personal development) and deviation from the normal quasi-experimental protocol. Thus, it was not possible to specifically split students in experimental and control groups whilst retaining the intent of the testing session and considering the autonomy.

Examples were administered, and the test methods were explained, including the use of multiple-choice answer sheet, demands of the assessment and use of examples for familiarity and understanding. Participants were informed about relevant time limit and provided with a clock to monitor timings. Completed answer sheets were collected and checked for clarity of response prior to optical scanning and passing through a software program. Electronic data scores were collated with first-test subject performance marks from the institute's management information systems. Data were anonymised and stored appropriately and securely for analysis.

### Data analyses

Data analyses were conducted on SPSS® version 25. Comparisons of the three time-limit groups were conducted using a one-way analysis of variance and Tukey's Honest Significant Difference (HSD) *post hoc* test of mean differences and significances. Pearson's  $r$  correlation coefficients and standard linear regression models (standardised beta weights because of range discrepancies) were used to assess the relationship between scores of tests and first-test marks.

## Ethical consideration

This study received ethical clearance from the Tshwane University of Technology Research Ethics Committee (No. REC/2016/09/001)

## Results

The Cloze procedure subtest yielded a maximum score of 20, whilst the Vocabulary in Context subtest score was out of a possible 30. First-test subject marks were expressed as a percentage value out of 100 possible points. Table 1 shows the mean values ( $M$ ) and standard deviations ( $SD$ ) of variables.

Table 1 shows similar levels of dispersion across different groups and subjects. Performance on the ELSA tests improved when time constraints were reduced but levels of dispersion remained stable despite differing sample sizes. No substantial differences in academic marks were present between the three time-limit groups.

### Differences between the time-limit groups

The one-way analysis of variance with Tukey's HSD *post hoc* revealed that the three time-limit groups differed significantly. The group without a time limit had higher scores on the Cloze procedure subtest ( $M = 13.93$ ,  $SD = 4.30$ ) than the double time-limit group ( $M = 12.22$ ,  $SD = 4.40$ ) or the normal time-limit group ( $M = 6.80$ ,  $SD = 4.08$ ). The one-way analysis of variance demonstrated that the groups differed significantly ( $F = 22.156$ ,  $p = 0.000$ ) and the Levene's test of homogeneity of variance met the required assumption of equal variances ( $F = 0.100$ ,  $p = 0.905$ ). The significant differences were identified as involving the normal time-limit group, for which scores were significantly lower than that of the double time-limit group ( $M_{\text{Difference}} = 5.442$ ,

$p = 0.000$ ) and the no time-limit group ( $M_{\text{Difference}} = 7.138$ ,  $p = 0.000$ ). However, the no time-limit and double time-limit groups did not differ significantly, despite slightly better performance by the no time-limit group ( $M_{\text{Difference}} = 1.716$ ,  $p = 0.440$ ). Similar findings were observed for the Vocabulary in Context subtest.

The no time-limit group performed best on the Vocabulary in Context subtest ( $M = 12.07$ ,  $SD = 4.98$ ), whilst the double time-limit group's scores were slightly lower ( $M = 11.13$ ,  $SD = 5.36$ ) and the normal time-limit group's scores were considerably lower ( $M = 6.48$ ,  $SD = 4.61$ ). The one-way analysis of variance revealed that the groups differed significantly ( $F = 10.902$ ,  $p = 0.000$ ) and the requirement of homogeneity of variance was satisfied ( $F = 0.666$ ,  $p = 0.517$ ). Examination of Tukey's HSD *post hoc* showed that the statistically significant differences were present between the normal time-limit group and the double time-limit group ( $M_{\text{Difference}} = 4.653$ ,  $p = 0.001$ ) as well as the no time-limit group ( $M_{\text{Difference}} = 5.589$ ,  $p = 0.001$ ). The no time-limit and double time-limit groups did not differ significantly ( $M_{\text{Difference}} = 0.936$ ,  $p = 0.833$ ). Therefore, significant differences were observed between the three time-limit groups, suggesting that time limitations influenced measuring English language skills by these tests. As a result, the timed conditions may also have affected the predictive power of each test.

### Prediction of first-test subject marks

Pearson's  $r$  correlation coefficients were calculated to examine the association between performance on the ELSA tests and performance in the first-test of each subject, followed by separate regression models for each group. Table 2 shows the correlation coefficients between the three time-limit mean values and subject performance.

**TABLE 1:** Descriptive statistics for the English Literacy Skills Assessment tests and first-test subject marks by time-limit group.

Assessment	All groups		Normal time limit		Double time limit		No time limit	
	Mean ( $M$ )	SD	Mean ( $M$ )	SD	Mean ( $M$ )	SD	Mean ( $M$ )	SD
Cloze procedure	9.62	5.19	6.80	4.08	12.22	4.40	13.93	4.30
Vocabulary in Context	8.80	5.46	6.48	4.61	11.13	5.36	12.07	4.98
Communications	52.38	13.63	51.80	15.19	54.57	11.04	50.73	12.76
Marketing for Tourism	70.13	20.37	68.80	20.60	70.13	21.26	74.07	19.10
Tourism Development	70.63	15.10	68.41	15.45	71.57	15.91	75.73	11.98
Travel and Tourism Management	63.36	18.80	65.00	17.08	60.17	22.25	63.20	18.59
Travel and Tourism Practice	59.95	11.80	60.11	11.30	57.39	11.82	63.40	13.02

$M$ , mean;  $SD$ , standard deviation.

**TABLE 2:** Pearson's  $r$  correlations between the English Literacy Skills Assessment tests and first-test subject marks by time-limit group.

Academic subject	Cloze procedure			Vocabulary in context		
	Normal time limit	Double time limit	No time limit	Normal time limit	Double time limit	No time limit
Communications	0.437**	0.473*	0.706**	0.313*	0.600**	0.634*
Marketing for Tourism	0.252	0.318	0.331	0.131	0.393	0.648**
Tourism Development	0.373*	0.359	0.574*	0.164	0.283	0.708**
Travel and Tourism Management	-0.049	0.104	-0.265	0.047	0.050	-0.138
Travel and Tourism Practice	0.450**	0.407	0.656**	0.265	0.544**	0.590*

\*, Significant at the 5% level.

\*\*, Significant at the 1% level.



Statistically significant positive correlations were present between the Cloze procedure subtest and the subject of 'Communications', which had a strong emphasis on English language. Similar coefficients were observed for the normal time-limit group ( $r = 0.437, p = 0.003$ ) and the double time-limit group ( $r = 0.473, p = 0.023$ ). A stronger statistically significant correlation was observed between the no time-limit group and the scores of the subject of 'Communications' ( $r = 0.706, p = 0.003$ ). The no time-limit group scores were also significantly correlated with scores of the first-test of tourism development ( $r = 0.574, p = 0.025$ ), whilst the normal time-limit group was less strongly, but more significantly, correlated ( $r = 0.373, p = 0.013$ ). The same is true about correlations between travel and tourism practice and Cloze procedure for the normal time-limit group ( $r = 0.450, p = 0.002$ ) and the no time-limit group ( $r = 0.656, p = 0.008$ ). No other statistically significant correlation coefficients were present. The correlational findings tentatively suggested that higher scores on the Cloze procedure test were associated with better performance on the subjects of 'Communications', 'Tourism Development' and 'Travel and Tourism Practice'. In most of the cases, the relationship between the scores and academic performance was strongest when no time limit was present, although the double time-limit coefficients were frequently similar. Significant positive correlation coefficients were also observed between the Vocabulary in Context test scores and the first-test subject marks, particularly if no time limit was implemented. Vocabulary in Context was more strongly associated with academic performance than the Cloze procedure.

Correlations between the subject of 'Communications' scores and Vocabulary in Context scores were statistically significant for the normal time-limit ( $r = 0.313, p = 0.038$ ), double time-limit ( $r = 0.600, p = 0.002$ ) and no time-limit groups ( $r = 0.634, p = 0.011$ ). The double time-limit group was also significantly correlated with 'Travel and Tourism Practice' scores ( $r = 0.544, p = 0.007$ ). However, only the no time-limit group was statistically significantly correlated with the first-test marks on 'Marketing for Tourism' ( $r = 0.648, p = 0.009$ ), 'Tourism Development' ( $r = 0.708, p = 0.003$ ) and 'Travel and Tourism Practice' ( $r = 0.590, p = 0.021$ ). For the Cloze procedure subtest, no statistically significant correlations were present with the first-test marks on 'Travel and Tourism Management'. For both tests, the no time-limit group appeared to be the most strongly associated group

with performance on the first-test of various subjects of tourism management, particularly the subject of 'Communications'.

Regression models were used to understand the relative predictive power of different time limit groups of each subject. Table 3 shows the standardised beta weights, statistically significant levels of the Cloze procedure subtest and coefficients of determination reporting the amount of variance explained.

When Cloze procedure is used as a predictor of the first-test marks, the regression on the subject of 'Communications' was strong, but the 'Marketing for Tourism' and 'Travel and Tourism Management' scores were not well predicted. Statistically significant increase in the SDs of first-test scores were associated with a single SD increase in Cloze procedure for the no time-limit group for the subjects of 'Communications' ( $\beta = 0.706, p = 0.003$ ), 'Tourism Development' ( $\beta = 0.574, p = 0.025$ ) and 'Travel and Tourism Practice' ( $\beta = 0.656, p = 0.008$ ). However, a slight inverse predictive function was observed for 'Travel and Tourism Management' ( $\beta = -0.265, p = 0.013$ ). The first-test scores for the subject of 'Communications' were also predicted by scores on the Cloze procedure for the normal time-limit group ( $\beta = 0.437, p = 0.003$ ) and the double time-limit condition ( $\beta = 0.473, p = 0.023$ ). The same was true for the subject of 'Travel and Tourism Practice' for the no time-limit ( $\beta = 0.656, p = 0.008$ ), double time-limit ( $\beta = 0.407, p = 0.054$ ) and normal time-limit groups ( $\beta = 0.450, p = 0.002$ ). For the subject of 'Travel and Tourism Practice', all three conditions had similar predictive power. For the Cloze procedure, no time limits resulted in stronger strength of prediction than doubling the time limits or implementing the normal time limit. Similar findings were present for the Vocabulary in Context test. The coefficients of determination, standardised regression values and probability values for Vocabulary in Context are shown in Table 4.

The Vocabulary in Context scores had statistically significant regression values for the subject of 'Communications' for the normal time-limit group ( $\beta = 0.313, p = 0.038$ ), double time-limit group ( $\beta = 0.600, p = 0.002$ ) and no time-limit group ( $\beta = 0.634, p = 0.011$ ). Standard deviation values of subjects were substantially increased with subtest increase for 'Travel and Tourism Practice' for both the double time-limit ( $\beta = 0.544, p = 0.007$ ) and the no time-limit groups ( $\beta = 0.590, p = 0.021$ ). However, the no time-limit group proved to be the strongest

**TABLE 3:** Regression values for the Cloze procedure test on first-test subject marks by time-limit group.

Academic subject	Normal time limit				Double time limit				No time limit			
	Constant	$R^2$	$\beta$	$p$	Constant	$R^2$	$\beta$	$p$	Constant	$R^2$	$\beta$	$p$
Communications	40.73	0.191	0.437	0.003	40.07	0.224	0.473	0.023	21.53	0.499	0.706	0.003
Marketing for Tourism	60.14	0.064	0.252	0.099	51.36	0.101	0.318	0.139	53.56	0.110	0.331	0.228
Tourism Development	58.81	0.139	0.373	0.013	55.70	0.129	0.359	0.092	53.47	0.329	0.574	0.025
Travel and Tourism Management	66.40	0.002	0.049	0.752	66.62	0.011	-0.104	0.635	79.17	0.139	-0.265	0.013
Travel and Tourism Practice	51.64	0.203	0.450	0.002	44.02	0.166	0.407	0.054	35.72	0.431	0.656	0.008

Note:  $R^2$  is the coefficient of determination representing the proportion of variance explained by the model.

**TABLE 4:** Regression values for the Vocabulary in Context test on the first-test subject marks by time-limit group.

Academic subject	Normal time limit				Double time limit				No time limit			
	Constant	R <sup>2</sup>	$\beta$	p	Constant	R <sup>2</sup>	$\beta$	p	Constant	R <sup>2</sup>	$\beta$	p
Communications	45.10	0.098	0.313	0.038	40.82	0.360	0.600	0.002	31.11	0.402	0.634	0.011
Marketing for Tourism	65.00	0.017	0.131	0.396	52.80	0.154	0.393	0.064	44.09	0.419	0.648	0.009
Tourism Development	64.85	0.027	0.164	0.287	62.24	0.080	0.283	0.191	55.17	0.502	0.708	0.003
Travel and Tourism Management	63.87	0.002	0.047	0.762	57.86	0.003	0.050	0.820	69.43	0.019	0.138	0.623
Travel and Tourism Practice	55.90	0.070	0.265	0.082	44.04	0.296	0.544	0.007	44.77	0.348	0.590	0.021

Note: R<sup>2</sup> is the coefficient of determination representing the proportion of variance explained by the model.

predictor, also having statistically significant power for the subjects of 'Marketing for Tourism' ( $\beta = 0.648, p = 0.009$ ) and 'Tourism Development' ( $\beta = 0.708, p = 0.003$ ). Like the Cloze procedure test, the regression of Vocabulary in Context test on the subject of 'Travel and Tourism Management' was poor and not statistically significant ( $p > 0.05$ ).

Both ELSA tests showed predictive power for the majority of the first-year subjects of tourism management based on statistically significant correlation coefficients and regression models. However, the no time-limit condition exhibited the strongest predictive power. Variance between ~33% and ~50% in academic first-test subject performance was explicable by English language proficiency measured on each of the two ELSA tests. In spite of not being significantly different from the no time-limit group, the double time-limit group did not show the same predictive relationship, potentially because of a truncated range of scores. The subject of 'Travel and Tourism Management', however, was not sufficiently associated with scores on either of the ELSA tests in terms of correlation or prediction.

## Discussion

The findings indicated that performance on both ELSA tests improved relatively to increase in time limitations. The statistics demonstrated that increased time limits resulted in a statistically significant improvement in performance, whilst the SD levels of mean values remained stable, suggesting that a consistent dispersion in scores was retained. Therefore, the findings reflected improvements in test outcome predictive quality when time limits are removed, despite the inherent limitations of comparing groups of differing sizes (Rosenthal & Rosnow, 2008). Nonetheless, similar improvements in English test outcomes were found by Hajebi et al. (2018) and Snow et al. (2009). In this regard, Harrington and Roche (2014) and Van der Linden (2011) also suggested that improvements in performance could be related to the more accurate assessment of constructs in the English language, rather than the ability to perform under time constraints. This disparity could be partially because of long-held notion of the influence of time constraints on the number of item responses and internal reliability of English proficiency tests themselves (Evans & Reilly, 1972).

Similar studies have suggested that implementing time constraints could reduce the reliability and validity of psychometric and language tests for a wide variety of

constructs (Lu & Sireci, 2007), resulting in the absence of equivalency across instruments (Cronbach & Warrington, 1951). Additionally, a biased presentation of English language ability is present if response levels below certain thresholds occur, or without readjustment of item functions (e.g. Van der Linden, 2011). The present research findings of improved performance without time constraints cannot necessarily be equated to changes in reliability or validity *per se* because of the absence of measurement of item response functions, despite studies such as those performed by Harrington and Roche (2014) being focused on similar assessment types. Nonetheless, Talento-Miller et al. (2013) also suggested that increasing the number of items attempted influenced the outcome of English language tests because of the varying difficulties and types of items rather than processing speed. The evidence suggests that inherent, internal test-structure issues under time-constrained conditions are influential, and the present findings concurred that working under time constraints could have negatively affected performance on both ELSA tests for this cohort. Although some other research has explored the inherent reliability issues surrounding time limits on English tests, the reviewed literature has not extensively explored the relative impact of differing time limits on predictive validity in the context of higher education. The regression analyses in the present research provided evidence of a predictive component for the two ELSA tests utilised, which strengthened when time limitations were extended or nulled.

The double time-limit and no time-limit groups' academic performance was positively and significantly correlated with performance on both ELSA tests, whilst the normal time-limit group demonstrated limited predictive power. Interestingly, predictive performance was similar for both double time-limit and no time-limit groups in most of the cases. This finding suggested that item response thresholds, such as those discussed by Van der Linden (2011) and Talento-Miller et al. (2013), could be important for predictive power as well as for internal consistency and reliability of measurement. Therefore, the present academic first-test performance could have been at least a partial function of English language ability, as measured by the ELSA tests. Several English language performance actions applied to the Cloze procedure protocol were used as one of the ELSA tests. However, in the present research, non-technical vocabulary levels measured in the context were found to be better predictors of academic performance.

Non-technical vocabulary levels have been successfully used as predictors in higher education institutions (HEIs) as well as a proxy for general English proficiency and Cloze procedures (Daller & Wang, 2014; Masrai & Milton, 2018; Qian, 2002; Schmitt et al., 2011; Snow et al., 2009; Trenkic & Warmington, 2018). The present study's findings suggest that vocabulary levels were more important in accurately predicting academic success than the Cloze procedure test, which required semantic manipulation and decision-making within the context of a passage. However, vocabulary ability could be subsumed into a variety of English functions present in the HEI performance requirements, such as lecture participation and development of text understanding and technical vocabulary. Vocabulary may be linked to other aspects of English language performance related to higher education, including deliberate performance and response selection (Macalister, 2010), improved heuristic learning of phrases and lexical translation (Koehn, Och, & Marcu, 2003), speed of translation and decoding within a finite memory capacity system (Sakurai, 2015), and meta-cognitive focus on syntactical awareness beside reformulation between languages in an attempt at better understanding (Jiménez et al., 2015). Known to be influenced by time constraints, some of these factors directly relate to essential skills measured in vocabulary and Cloze procedure tests, including semantic representations, understanding of words in context, reading speed and quality and the ability to manipulate syntactical arguments.

Reported findings that English proficiency tests encompassing vocabulary, grammar and contextual representation are affected by time limitations (e.g. Bridgeman et al., 2004; Murray, 2010) were confirmed by the present research using two ELSA tests. Such content-specific skill development for understanding may require measurement outside of what could generally be considered as the normal, time-constrained and psychometrically focused framework. Furthermore, development of content/subject-specific technical language could also play a role in academic outcomes, particularly if basic levels have not been fully developed as a foundation (Birrell, 2006). The findings of the present research also suggest that time limitations play an important role in performance and predictive validity, beside choice of test for predictive purposes. Removal of time limitations resulted in more accurate prediction of academic success outcomes, and use of the Vocabulary in Context test resulted in the strongest predictive power. These findings suggested that appropriate English proficiency assessment could hinge more on the determination of specific academic weaknesses within English language whilst reducing the role of time limitation as an essential factor in predicting performance. In spite of the various findings suggesting that time constraints impact a variety of factors concerning English proficiency tests, from a practical perspective, it is unlikely that performing lengthy tests without time limits would be practical in the context of real world. Nonetheless, studies suggest that time constraints could alter the psychometric properties of tests in a variety of ways.

In spite of important findings in the present research, the study carried some limitations which created some uncertainty in the interpretation of the results. Groups of unequal sizes, because of the voluntary nature of participation, may have resulted in misrepresentation of values because of the use of parametric statistics in such a case (Rosenthal & Rosnow, 2008). Similarly, small groups and lack of randomisation could have affected the statistical outcomes. An example of this issue could be the negative correlations seen in for the subject of 'Travel and Tourism Management', although alternate explanations such as subject content could also account for this anomaly. Nonetheless, inequality in ranges of scores between different variables still resulted in Pearson's  $r$  and a linear regression model being the most suitable choice, albeit imperfect. In addition, it was not possible to fully standardise the English language pre-entry (Grade 12) performance in this case. Therefore, this criterion was only passively standardised as a minimum level through the use of a specific qualification grouping of students. Pre-entry English ability could have impacted the outcome on either of the English proficiency tests, thus introducing bias in the results or impacting the selection of groups in an attempt by the participants to maximise their performance. Nonetheless, it is believed that present language ability, regardless of prior ability, is the most important factor in interpreting the findings, because the intention is to predict academic performance rather than investigate validity of the assessments in question. Furthermore, the results appear to indicate that time limitations imposed on English proficiency tests are of importance in fully applying the concept of language proficiency to higher education outcomes.

## Conclusion

The present research findings demonstrate that performance and predictive power on the modified ELSA versions of Cloze procedure and Vocabulary in Context improves when time limits are increased or removed. The findings imply that factors such as item completion thresholds, reading speed, semantic understanding, and translation for decision-making requirements could contribute to negative changes in performance under time-constrained conditions. Therefore, students may possess some of the English language skills associated with academic performance but are unable to demonstrate these skills within the imposed time constraints. Although these findings are useful, they should be treated with caution as current internal reliability and predictive validity data are not available for full assessment and this pilot study was conducted on smaller, unequal sample groups. Nonetheless, it is apparent that the English proficiency as measured by the ELSA could be inaccurately reflected under time-constrained conditions, limiting the ability of the test to serve as a predictor of academic performance in tertiary education. These findings imply that further investigations are required to develop sufficiently



competency gap-targeted English interventions, and the future research should consider larger-scale studies to identify specific components within the tests which contribute to academic success in South African HEIs.

## Acknowledgements

### Competing interests

The author has declared that no competing interest exists.

### Author's contributions

I declare that I am the sole author of this research article.

### Funding information

The research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

### Data availability statement

Data sharing is negotiable by request. Sharing of data cannot be guaranteed and will depend on the nature of the request.

### Disclaimer

The views and opinions expressed in this article are those of the author and do not necessarily reflect the official policy or position of any affiliated agency of the author.

## References

- Abriam-Yago, K., Yoder, M., & Kataoka-Yahiro, M. (1999). The Cummins model: A framework for teaching nursing students for whom English is a second language. *Journal of Transcultural Nursing*, 10(2), 143–149. <https://doi.org/10.1177/104365969901000208>
- Anderson, N.J. (1991). Individual differences in strategy use in second language reading and testing. *The Modern Language Journal*, 75(4), 460–472. <https://doi.org/10.1111/j.1540-4781.1991.tb05384.x>
- Andrade, M.S. (2006). International students in English-speaking universities. *Journal of Research in International Education*, 5(2), 131–154. <https://doi.org/10.1177/1475240906065589>
- Arrigoni, E., & Clark, V. (2015). *Investigating the appropriateness of IELTS cut-off scores for admissions and placement decisions at an English-medium university in Egypt*. IELTS Research Report Series. Retrieved from <https://www.ielts.org/teaching-and-research/research-reports>
- Bedenlier, S., & Zawacki-Richter, O. (2015). Internationalization of higher education and the impacts on academic faculty members. *Research in Comparative & International Education*, 10(2), 185–201. <https://doi.org/10.1177/174549915571707>
- Benzie, H.J. (2010). Graduating as a 'native speaker': International students and English language proficiency in higher education. *Higher Education Research & Development*, 29(4), 447–459. <https://doi.org/10.1080/07294361003598824>
- Birrell, B. (2006). Implication of low English standards among overseas students at Australian universities. *People and Place*, 14(4), 53–64.
- Bridgeman, B., McBride, A., & Monaghan, W. (2004). *Testing and time-limits*. Princeton, NJ: Educational Testing Services.
- Bruton, C., Wisessuwan, A., & Tubsree, C. (2018). Praxial interlanguage experience: Developing communicative intentionality through experiential and contemplative inquiry in international education. *HRD Journal*, 9(1), 27–36.
- Casale, D., & Posel, D. (2011). English language proficiency and earnings in a developing country: The case of South Africa. *Journal of Behavioral and Experimental Economics* 40(4), 385–393.
- Coleman, J.A. (2006). English-medium teaching in European higher education. *Language Teaching*, 39(1), 1–14. <https://doi.org/10.1017/S026144480600320X>
- Cronbach, L.J., & Warrington, W.G. (1951). Time-limit tests: Estimating their reliability and degree of speeding. *Psychometrika*, 16(2), 167–188. <https://doi.org/10.1007/BF02289113>
- Cross, M., & Carpentier, C. (2009). 'New students' in South African higher education: Institutional culture, student performance and the challenge of democratisation. *Perspectives in Education*, 27(1), 6–18.
- Cummins, J. (2000). *Language, power, and pedagogy: Bilingual children in the crossfire*. New York, NY: Multilingual Matters. ISBN: 9781853594731
- Daller, M., & Wang, Y. (2017). Predicting study success of international students. *Applied Linguistics Review*, 8(4), 355–374. <https://doi.org/10.1515/applirev-20162013>
- Dalton-Puffer, C. (2011). Content-and-language integrated learning: From practice to principles? *Annual Review of Applied Linguistics*, 31, 182–204. <https://doi.org/10.1017/S0267190511000092>
- Daly, J.L., & Stahmann, R.F. (1968). The effect of time-limits on a university placement test. *The Journal of Educational Research*, 62(3), 103–104. <https://doi.org/10.1080/00220671.1968.10883779>
- Doe, C. (2014). Diagnostic English language needs assessment. *Language Testing*, 31(4), 537–543. <https://doi.org/10.1177/0265532214538225>
- Escamilla, K. (2009). English language learners: Developing literacy in second-language learners – Report of the national literacy panel on language-minority children and youth (D. August, & T. Shanahan, Eds.). *Journal of Literacy Research*, 41, 432–452. <https://doi.org/10.1080/10862960903340165>
- Evans, F., & Reilly, R. (1972). A study of speededness as a source of test bias. *Journal of Educational Measurement*, 9(2), 123–131. <https://doi.org/10.1111/j.1745-3984.1972.tb00767.x>
- Fairbairn, S. (2007). Facilitating greater test success for English language learners. *Practical Assessment, Research & Evaluation*, 12(11), 1–7.
- Feast, V. (2002). The impact of IELTS scores on performance at university. *International Education Journal*, 3(4), 70–85.
- Fenton-Smith, B., Humphreys, P., & Walkinshaw, I. (2018). On evaluating the effectiveness of university-wide credit-bearing English language enhancement courses. *Journal of English for Academic Purposes*, 31, 72–83. <https://doi.org/10.1111/j.1745-3984.1972.tb00767.x>
- Gellert, A.S., & Elbro, C. (2013). Cloze tests may be quick, but are they dirty? Development and preliminary validation of a Cloze test of reading comprehension. *Journal of Psychoeducational Assessment*, 31(1), 16–28. <https://doi.org/10.1177/0734282912451971>
- Goto, K., Maki, H., & Kasai, C. (2010). The minimal English test: A new method to measure English as a second language proficiency. *Evaluation & Research in Education*, 23(2), 91–104. <https://doi.org/10.1080/09500791003734670>
- Hajebi, M., Taheri, S.Q., & Allami, H. (2018). A comparative study of cloze test and C-test in assessing collocational competence of Iranian EFL learners. *European Online Journal of Natural and Social Sciences*, 7(1), 225–234.
- Harrington, M., & Roche, T. (2014). Identifying academically at-risk students at a English-as-a-Lingua-Franca university setting. *Journal of English for Academic Purposes*, 15, 37–47. <https://doi.org/10.1016/j.jeap.2014.05.003>
- Huetting, F. (2015). Four central questions about prediction in language processing. *Brain Research*, 1626, 118–135. <https://doi.org/10.1016/j.brainres.2015.02.014>
- Jackson, D. (2015). Employability skill development in work-integrated learning: Barriers and best practice. *Studies in Higher Education*, 40(2), 350–367. <https://doi.org/10.1080/03075079.2013.842221>
- Jiménez, R.T., David, S., Fagan, K., Risko, V.J., Pacheco, M., Pray, L., et al. (2015). Using translation to drive conceptual development for students becoming literate in English as an additional language. *Research in the Teaching of English*, 49, 248–271.
- Kaleidoprax. (2014). What ELSA measures? Retrieved n.d. from <https://www.kaleidoprax.co.za/english-literacy-skills-assessment.html>
- Keenan, J.M., Betjemann, R.S., & Olson, R.K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, 12(3), 281–300. <https://doi.org/10.1080/10888430802132279>
- Koehn, P., Och, F.J., & Marcu, D. (2003). Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology (NAACL '03)* (vol. 1, pp. 48–54). Edmonton: Association for Computational Linguistics. <https://doi.org/10.3115/1073445.1073462>
- Lu, Y., & Sireci, S.G. (2007). Validity issues in test speededness. *Educational Measurement: Issues and Practice*, 26(4), 29–37. <https://doi.org/10.1111/j.1745-3992.2007.00106.x>
- Luke, S.G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, 88, 22–60. <https://doi.org/10.1016/j.cogpsych.2016.06.002>
- Macalister, J. (2010). Investigating teacher attitudes to extensive reading practices in higher education: Why isn't everyone doing it? *RELJ Journal*, 41(1), 59–75.
- MacIntyre, P.D., & Gardner, R.C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning*, 44(2), 283–305.
- Masrai, A., & Milton, J. (2018). Measuring the contribution of academic and general vocabulary knowledge to learners' academic achievement. *Journal of English for Academic Purposes*, 31, 44–57. <https://doi.org/10.1177/0033688210362609>
- Millin, T., & Millin, M. (2018). English academic writing convergence for academically weaker senior secondary school students: Possibility or pipe-dream? *Journal of English for Academic Purposes*, 31, 1–17. <https://doi.org/10.1016/j.jeap.2017.12.002>
- Murray, N.L. (2010). Conceptualising the English language needs of first year university students. *The International Journal of the First Year in Higher Education*, 1(1), 55–64. <https://doi.org/10.5204/intjfyhe.v1i1.19>
- Nunan, D. (2003). The impact of English as a global language on educational policies and practices in the Asia-Pacific region. *TESOL Quarterly*, 37(4), 589–613. <https://doi.org/10.5204/intjfyhe.v1i1.19>



- Powers, D.E., & Fowles, M.E. (1997). *Effects of applying different time-limits to a proposed GRE writing test*. Princeton, NJ: Educational Testing Service.
- Prinsloo, C.H., & Heugh, K. (2013). *The role of language and literacy in preparing South African learners for educational success: Lessons learnt from a classroom study in Limpopo province*. Pretoria: Human Sciences Research Council.
- Qian, D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, 52(3), 513–536. <https://doi.org/10.1111/1467-9922.00193>
- Read, J. (2008). Identifying academic language needs through diagnostic assessment. *Journal of English for Academic Purposes*, 7(3), 180–190. <https://doi.org/10.1016/j.jeap.2008.02.001>
- Rosenthal, R., & Rosnow, R.L. (2008). *Essentials of behavioral research* (3rd edn.). New York, NY: McGraw Hill.
- Sakurai, N. (2015). The influence of translation on reading amount, proficiency, and speed in extensive reading. *Reading in a Foreign Language*, 27(1), 96–112. ISSN 1539-0578.
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, 95, 26–43. <https://doi.org/10.1111/j.1540-4781.2011.01146.x>
- Snow, C.E., Lawrence, J.F., & White, C. (2009). Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness*, 2(4), 325–344. <https://doi.org/10.1080/19345740.903167042>
- Solano-Flores, G. (2008). Who is given tests in what language by whom, when and where? The need for probabilistic views of language in the testing of English language learners. *Educational Researcher*, 37(4), 189–199. <https://doi.org/10.3102/0013189X08319569>
- Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language*, 82, 1–17. <https://doi.org/10.3102/0013189X08319569>
- Sun, C., & Henrichsen, L. (2010). Major university English tests in China: Their importance, nature and development. *TESL Reporter*, 44, 1–24.
- Talento-Miller, E., Guo, F., & Han, K.T. (2013). Examining test speededness by native language. *International Journal of Testing*, 13, 89–104. <https://doi.org/10.1080/15305058.2011.653021>
- Taylor, S., & von Fintel, M. (2016). Estimating the impact of language instruction in South African primary schools: A fixed effects approach. *Economics of Education Review*, 50, 75–89. <https://doi.org/10.1016/j.econedurev.2016.01.003>
- Tomasello, M. (2014). *A natural history of human thinking*. Boston, MA: Harvard University Press. ISBN: 9780674724778
- Trace, J., Brown, J.D., Janssen, G., & Kozhevnikova, L. (2017). Determining cloze item difficulty from item passage characteristics across different learner backgrounds. *Language Testing*, 34(2), 151–174. <https://doi.org/10.1177/0265532215623581>
- Trenkic, D., & Warmington, M. (2018). Language and literacy skills of home and international university students: How different are they, and does it matter? *Bilingualism: Language and Cognition*, 22(2), 349–365. <https://doi.org/10.1017/S136672891700075X>
- Van der Linden, W.J. (2011). Test design and speededness. *Journal of Educational Measurement*, 48(1), 44–60. <https://doi.org/10.1111/j.1745-3984.2010.00130.x>
- Webb, V. (2002). English as a second language in South Africa's tertiary institutions: A case study at the University of Pretoria. *World Englishes*, 21(1), 49–61. <https://doi.org/10.1111/1467-971X.00231>