


The impact of different time limits and test versions on reliability in South Africa

**Author:**Danille E. Arendse¹ **Affiliation:**

¹Research Consultancy Department, Military Psychological Institute, Pretoria, South Africa

Corresponding author:

Danille Arendse,
arendse.danielle@gmail.com

Dates:

Received: 18 Mar. 2019

Accepted: 10 Jan. 2020

Published: 03 Mar. 2020

How to cite this article:

Arendse, D.E. (2020). The impact of different time limits and test versions on reliability in South Africa. *African Journal of Psychological Assessment*, 2(0), a14. <https://doi.org/10.4102/ajopa.v2i0.14>

Copyright:

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

The empirically developed English comprehension test (ECT) was created for organisational and educational purposes to assess verbal reasoning. The initial version of the ECT had an associated time limit of 45 min, which required individuals to complete it within the specified time, while the later version of the ECT had no time limit. The ECT's two test versions – a timed and an untimed version – were piloted as part of the development and validation of the ECT. The purpose of this article was to explore the internal consistency of the two test versions and compare the reliability of the timed and untimed versions of the ECT. This study was conducted to establish whether reliability was affected by the different time limit-related requirements. The sample size for ECT version 1.2 was 597 and ECT version 1.3 comprised 882 individuals. The methods used for comparison in this article involved a graphical display of performance relating to both test versions and an exploration of the times recorded for the untimed test version. A reliability analysis was performed to evaluate the internal consistency of the two test versions. The performance of individuals in the untimed and timed versions of the ECT was similar based on the average minimum and maximum scores. The Cronbach's alpha indicated that verbal reasoning was measured consistently for the two test versions. This result suggested that time did not negatively affect the reliability of the test.

Keywords: psychometrics; reading comprehension; reliability; Cronbach's alpha; test performance; timed assessment; time limit.

Introduction

Reading comprehension comprises cognitive and linguistic components that support an individual in generating meaning from a text. The core processes that assist in producing meaning from texts are decoding and comprehension (Kendeou, Van den Broek, Helder, & Karlsson, 2014; Pretorius, 2002). These processes of decoding and comprehension are inter-related and associated with reading and literacy (Kannianen, Killi, Tolvanen, Aro, & Leppanen, 2019). Literacy is an important aspect of learning and is required in various facets of one's life. It begins with early schooling and continues until necessary as part of executing one's job or for studying further. Because literacy involves reading and making meaning, it was of concern that South African research about grades 3 and 4 learners indicated that they were struggling to comprehend and derive meaning from texts at school (Howie et al., 2017; Spaull, 2016). Although these studies refer to the literacy levels of grades 3 and 4 learners, the reality of a literacy crisis in South Africa strikes when learners' foundations in literacy and reading are not in place, which is also one of the reasons for the low retention of learners until they complete Grade 12 (Spaull, 2013). The significance of literacy skills is that it affects reading and the performance of learners in reading comprehension and verbal assessments (Kannianen et al., 2019). Moreover, reading comprehension is a component of learning English (Bahardoost & Ahmadi, 2018). The literacy and comprehension levels of English additional language learners (this refers to learners who are non-native English speakers) can, however, be affected by the availability of resources and the socio-economic status (SES) of these learners, factors that negatively affect the quality of the education they receive (Cockcroft, Bloch, & Moolla, 2016; Spaull, 2016). The educators' level of competence in English, which causes them to use code switching (this refers to the mixture of English and other native languages in South Africa), is also a factor contributing to low English literacy amongst non-native English speakers (Krugal & Fourie, 2014; Kuwornu, 2017). In a context other than that of South Africa, the Australian context has, for example, identified similar issues related to English language comprehension and literacy faced by individuals from an Aboriginal background (Dingwall, Gray, McCarthy, Belima, & Bowden, 2017; Dingwall, Lindeman, & Cairney, 2014). A consideration of the factors that affect literacy and comprehension is important when evaluating an English test in South Africa.

Read online:

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

The English comprehension test (ECT) is theorised to measure verbal reasoning. Furthermore, the ECT is a South African test initiative that addresses the need to develop local tests that provide for the multicultural context in which tests in South Africa are used (Bekwa, 2016). The development and refinement of the test, which are still underway, have led to two test versions (ECT version 1.2 and version 1.3) so far. The reasoning behind the two test versions was that the latter (ECT 1.3) would be an improved version of the former (ECT 1.2). The removal of a time limit is one of the changes that was made with respect to both test versions and is the specific focus of this article. Therefore, the compromise between speed and ability is a significant factor when evaluating the reliability of the ECT (Goldhammer, 2015; Streiner, 2003). The adjustment of test conditions of assessments, such as extending the time allocated for the assessment, can be likened to a process of accommodation. Accommodation in relation to the ECT can be viewed as a form of support that allows test-takers to show their understanding of the assessment (Kuwornu, 2017).

Timed assessments, particularly in the case of power tests, may affect reliability because they are focused on the items completed rather than on the responses to items (Goldhammer, 2015; Lee & Chen, 2011; Streiner, 2003). It is thus imperative to explore the impact of time on the reliability of the ECT and to investigate the actual time required by the slowest person to complete the test. Although the slowest person could be an outlier, the focus was on allowing all the participants to complete the test, thereby fully accommodating individuals in the assessment. Thus, the need to extend time limits in a multicultural context such as South Africa is an imperative consideration for item completion. There are within-learner factors such as reading speed and coding or de-coding processes while reading, which are also worth noting (Goldhammer, 2015; Kendeou et al., 2014; Pretorius, 2002). This also relates to the intention of the ECT, which is primarily focused on eliciting the ability to decode texts and not on the ability under time-related pressure. The influence of time limits may also cause the working memory of the individual to be measured instead of the intended construct (Keith & Reynolds, 2010; Oberauer & Lewandowsky, 2013).

The importance of being able to read and infer from text as well as to create meaning from text is implied in the ECT (Arendse & Maree, 2019). In a study on the factor structure of the ECT (Arendse & Maree, 2019), it was also indicated that the ECT has a definite cognitive component. Furthermore, the factors emerging from the ECT, namely, reasoning, deduction and vocabulary, are directly related to reading and comprehension. The factors were labelled based on the content of their loadings, and the outcome suggested that the ECT was possibly a measure of cognitive (verbal) ability. This commonality of factors and cognitive (verbal) ability was found across the two test versions. It was also argued that ECT version 1.3 had a theoretically stronger factor structure, thereby suggesting that ECT 1.3 was an improved test version (Arendse & Maree, 2019). The results of the study indicated that there was a definite dominant factor that emerged from

both test versions. The evidence of the dominant factor was, however, not sufficient to claim that the test was unidimensional (Arendse & Maree, 2019). This is an important consideration for exploring the reliability of the ECT, as the Cronbach's alpha is sensitive to multidimensional scales (Abedi, 2002; Osburn, 2000; Streiner, 2003; Taber, 2018; Tavakol & Dennick, 2011). The cognitively influenced factors (i.e. reasoning, deduction and vocabulary) of the ECT are also affected by the reliability of the ECT, as processes associated with reading comprehension and reasoning may be hampered. Thus, time limits may have an influence on individuals' reasoning ability and reading processes (reading speed and decoding).

Considering the findings related to literacy and comprehension levels in South Africa, individuals who are not English first-language speakers may have some difficulty in completing English assessments (Van de Vijver & Rothmann, 2004). Thus, the addition of an imposed time limit could affect individuals' true reflection of ability in assessments (Angelidis, Solis, Lautenbach, Van der Does, & Putman, 2019). The aspects related to timed assessments are important to acknowledge when considering the two versions of the ECT.

The rationale for this study was to establish whether any differences were observed in the reliability of the two versions of the ECT when different time limits are applicable, one being the lack of a time limit. Although there were substantial differences between the two test versions, the majority of the items remained the same. These changes across test versions are discussed in the 'Instrument' section of this article. Because time may play a role in performance in tests, it is essential to explore the reliability of the two test versions. These findings will provide important insights regarding the effects of time limits on reliability.

The objectives of the study were as follows:

Objective 1: to assess how long individuals were taking to complete the ECT by exploring the recorded times of the ECT 1.3

Objective 2: to explore the internal reliability of the two test versions of the ECT using Cronbach's alpha.

Methods

This study was quantitative in nature as the aims of this article were aligned with quantitative data analysis. The assessment of recorded times was done by physically recording the time that the last person completed the ECT, which was thereafter captured using Microsoft Excel. In addition, the average of the recorded times was also interpreted. The assessment of the reliability of the two test versions involved the use of Cronbach's alpha, which was calculated using the Statistical package for the Social Sciences (SPSS) 23 package.

Being the reliability coefficient, Cronbach's alpha is commonly used in psychology to assess internal consistency and was therefore used in this study (Cronbach, 1951; Osburn, 2000). Because each test version of the ECT was administered once,

Cronbach's alpha was suitable for measuring reliability (Tavakol & Dennick, 2011). Cronbach's alpha was also used because the other reliability statistics measurements, such as Guttman's Lambda 4 (Guttman, 1945) is calculated using a split-half method rather than internal correlations, thus analysing the internal covariance between the two halves. It was also considered problematic that Guttman's Lambda 4 (Guttman, 1945) demands more stringent requirements regarding the sample size and length of test; thus, Cronbach's alpha was deemed more suitable (Abedi, 2002; Cronbach, 1951; Erguven, 2014; Osburn, 2000; Streiner, 2003; Taber, 2018; Tavakol & Dennick, 2011).

Participants

The study sample comprised 597 individuals for test version 1.2 and 881 individuals for test version 1.3, respectively. For ECT version 1.2, respondents' ages ranged from 18 to 52 years (mean age = 22 years) and for ECT version 1.3, the age groups ranged from 18 to 42 years (mean age = 21 years). The individuals completing the ECT 1.2 were, however, predominately under the age of 35 years ($n = 572$, 96%), with only 24 (4%) individuals over the age of 35. The individuals completing the ECT 1.3 were mainly under the age of 35 ($n = 786$, 89%), while 6 (1%) individuals were 36 years old and older. All nine provinces and 11 languages in South Africa were represented in the sample for both test versions (Arendse, 2018; Arendse & Maree, 2019). Table 1 presents participants' demographics for the two test versions. In the table, the percentage of gender representation is indicated, which was predominantly male in both test versions. The distribution of home languages is also indicated in terms of the following categories: English, Afrikaans and African languages. This language distribution is significant for the interpretation of the reliability of the ECT across the two test versions. Moreover, the majority of the samples are non-native English speakers (Afrikaans and African languages).

Data collection instruments

The ECT is an individual test that is theorised to assess an individual's verbal reasoning ability (Arendse, 2018; Arendse & Maree, 2019). The ECT contains a comprehension section that is

TABLE 1: Differences between the two test versions of the English comprehension test.

Participant demographics	ECT 1.2	ECT 1.3
Gender	66% males 27% females 7% missing	75% males 24% females 1% missing
Age	18–52 years	18–42 years
Home languages	7% Xhosa 10% Zulu 14% Sepedi 18% Setswana 6% Venda 5% Tsonga 1% Ndebele 7% Sotho 2% Siswati 15% Afrikaans 7% English	7% Xhosa 11% Zulu 16% Sepedi 13% Setswana 6% Venda 6% Tsonga 3% Ndebele 9% Sotho 3% Siswati 15% Afrikaans 8% English

Source: Arendse, D.E. (2018). *Exploring the construct validity and reliability of the English comprehension test*. (Unpublished doctoral thesis). Pretoria: University of Pretoria; and, Arendse, D.E., & Maree, D. (2019). Exploring the factor structure of the English comprehension test. *South African Journal of Psychology*, 49(3), 376–390. <https://doi.org/10.1177/0081246318805268>

ECT, English comprehension test.

made up of multiple-choice questions. The language section contains multiple-choice questions that have four answer options, with only one option indicating the correct answer, and a written answer section (sentence construction items). The scoring for both the comprehension and language sections was dichotomous. In Table 2, an example of a comprehension question from ECT versions 1.2 and 1.3 is presented.

Although the test is still under development, it was administered to individuals from different linguistic and cultural backgrounds in South Africa. The ECT has only been used for research purposes and thus the initial test version, ECT 1.2, was essentially a pilot research study. A preliminary item analysis of ECT version 1.2 indicated that there were some problematic items and for this reason one problematic item was edited and two items were removed. Although there are substantial differences between the administrations of the two test versions, the majority of the items across the two test versions remained the same. English comprehension test version 1.3 included five new items (plurals), such as foot or feet, to assess other language aspects not previously covered in ECT version 1.2. Table 3 presents an example of a test question from ECT version 1.3.

The age groups for the pilot study of the ECT were between 18 and 52 years. This broad age group was tested following the convenience sampling method and a maximum sample was retained. The range of the age group should, however, be viewed with caution as the majority of the individuals participating across both test versions were actually under 25 years of age.

English comprehension test version 1.3, predominately based on the content of ECT 1.2, with some changes (indicated in Table 4), was used for research purposes. English comprehension test version 1.2 has 39 items and a time limit of 45 min was imposed. English comprehension test version 1.3 has 42 items and no time limit was imposed (Arendse, 2018; Arendse & Maree, 2019). In Table 4, the changes made across the test versions are indicated. These changes across test versions are worth noting as possible factors that may have had an impact on the reliability of the ECT.

TABLE 2: Example of a test question (comprehension) in the English comprehension test versions 1.2 and 1.3.

Option	Description
Which statement is TRUE according to the information given in the passage?	
A.	The passengers got out of the aircraft because they wanted to see the back of the aircraft during refuelling.
B.	The passengers got out of the aircraft because they wanted to stand by the flight line during refuelling.
C.	The passengers got out of the aircraft because they could not be in the aircraft during refuelling.
D.	The passengers got out of the aircraft because they felt odd and nervous during refuelling.

TABLE 3: Example of a test question (plurals) in the English comprehension test version 1.3.

Option	Description
Choose the correct form of the word in the following sentences.	
1.	The navy men dived into the sea and injured their against the hidden rocks. (A. foot, B. feet)

TABLE 4: Differences between the two versions of the English comprehension test.

Changes across versions of the ECT	ECT 1.2	ECT 1.3
Instructions	Limited instructions and sections of the test were demarcated in a limited manner	More detailed instructions and the different sections of the test were demarcated more clearly
Total items	39 items	42 items
Time limit	45 min	No time limit, but the time when the last person completed the test was recorded

Source: Arendse, D.E. (2018). *Exploring the construct validity and reliability of the English comprehension test*. (Unpublished doctoral thesis). Pretoria: University of Pretoria; and, Arendse, D.E., & Maree, D. (2019). Exploring the factor structure of the English comprehension test. *South African Journal of Psychology*, 49(3), 376–390. <https://doi.org/10.1177/0081246318805268>

ECT, English comprehension test.

Procedure

The sampling method used to obtain the data was convenience sampling, as the individuals in the study were all attending selections, thus making them accessible for the ECT pilot. The individuals were attending selection sessions for possible employment and were available after the selection process had been completed owing to the transport arrangements that had been made for them. The ethics of exploring the test for further validation after a high-stakes process had completed was considered and the individuals were informed that the test was for research purposes and thus were not compelled to participate in the research process. The ECT is intended for screening purposes and not high-stakes testing. Furthermore, the accessibility of the sample allowed for the piloting of the ECT. The reasoning behind piloting the ECT after lunch was that the research should not affect the performance of individuals in the selection process relevant to the employment for which they were applying. The performance of the participants in the research after the selection may have been affected by some measure of stress caused by the selection, which could either have inhibited or enhanced their performance. All the candidates were either Grade 12 learners or had already completed Grade 12. Before the selection began, the participants completed the informed consent document. The participants were informed of the ECT and asked whether they would consent to taking the test for research purposes. After the selection process had been concluded, the participants were given a lunch break and thereafter they completed the ECT.

The time of day that the research testing took place was early afternoon, which could imply that several factors may have had an impact on their performance in the ECT. These factors include stress, fatigue, attitude, motivation and the energy levels of the participants when completing an assessment (Angelidis et al., 2019; Bunyi et al., 2015; Dingwall et al., 2017; Dodeen, Abdelfattah, & Alshumrani, 2014; Kiwan, Ahmed, & Pollitt, 2000; Kuwornu, 2017). Although some individuals may employ stress as a motivator, others may experience stress as inhibiting their performance and causing anxiety (Bunyi et al., 2015). These factors need to be acknowledged as they may have had an influence on the individuals completing both the untimed and timed versions of the ECT.

The administration of these pilot sessions involved test orientation and assisting individuals with completing the biographical section of the answer sheet (Arendse, 2018; Arendse & Maree, 2019). The test times for each session were recorded manually, while only the starting time of the test and the time that the last person completed the test were recorded. The time recorded in the latter instance was therefore based on the maximum time required for the slowest person to complete the test. The reason for doing this was to assess the maximum time taken by an individual to complete the ECT. A serious limitation with this method of recording time was that an average completion time could not be calculated.

The ethical considerations were appropriately applied in this study. The confidentiality and privacy of participants were respected, with a view to keep any identifying information private and confidential. The participants, as said before, signed the informed consent document, which is a standard practice and allowed the individuals to be informed of what the research entails and that they were not forced to participate. The safeguarding of information is important and all data have been put into safekeeping. The data may only be accessed by registered professionals. Ethical clearance for this study was obtained from the University of Pretoria (Arendse, 2018; Arendse & Maree, 2019).

Data analysis

The description of the data includes the observation of skewness and kurtosis statistics to assess the normality of the data. This was done using SPSS employing descriptive statistics. The performance of individuals across the two test versions was indicated by means of scatter plots, which were generated using Microsoft Excel. The performance in the test involved a representation of the number of correct and incorrect answers to the questions in the test. This representation was done for both versions of the ECT. It should be noted that the missing data for ECT 1.3 were not captured because of the scanning process that automatically scored the test. For this reason, the incorrect and missing data would have been captured similarly (as a 0) for ECT 1.3. Because the missing data could not be compared across test versions, it was not included in the scatter plot. Moreover, the incorrect data are potentially inflated because of possible missing data, which therefore presents a limitation to this study. Although one would not expect missing data when no time limit is imposed, individuals completing the test were not forced to complete all test items.

The reliability coefficient, Cronbach's alpha, was calculated using SPSS and used to assess how consistent the items of the test were as a whole (Cronbach, 1951; Hedge, Powell, & Sumner, 2018; Liao, 2004; Santos, 1999; Streiner, 2003; Taber, 2018; Tavakol & Dennick, 2011). The Cronbach's alpha contains information about how correlated the items of the test are to one another, which is referred to as the internal consistency of the measure. The Cronbach's alpha associated with the reliability in examining the internal consistency of the scale ranges from 0 to 1; thus, the closer this value is to 1, the more reliable the test will be in measuring the construct (Mushquash & Bova, 2007; Streiner, 2003; Taber, 2018; Tavakol & Dennick, 2011).

Ethical consideration

Ethical clearance to conduct the study was obtained from the University of Pretoria (GW20150407HS).

Results

The descriptive statistics in ECT version 1.2 indicated a range of scores from 8 to 38, with an average score of 23. The descriptive statistics in ECT version 1.3 indicated a range from 8 to 39, with an average score of 26. It is important to inspect the symmetry of the data to justify the use of parametric analyses across the two test versions. The skewness of -0.125 and the kurtosis of -0.284 for ECT version 1.2 indicate that the data are fairly symmetrical and have a flat distribution (Field, 2009). These values are within the commonly accepted range of -1.000 to +1.000. The Kolmogorov–Smirnov and Shapiro–Wilks tests for ECT version 1.2 were, respectively, $D(597) = 0.055$, $p < 0.05$ and $D(597) = 0.994$, $p < 0.001$, indicating that the data are significantly non-normal (Field, 2009).

The skewness of -0.256 and the kurtosis of -0.082 for ECT version 1.3 indicate that the data are slightly negatively skewed and have a flat distribution (Field, 2009). This suggests that the majority of responses fell towards or above the mean value. The Kolmogorov–Smirnov and Shapiro–Wilks tests for ECT version 1.3 were, respectively, $D(881) = 0.063$, $p < 0.001$ and $D(881) = 0.987$, $p < 0.001$, which indicates that the data are significantly non-normal (Field, 2009).

According to the skewness and kurtosis values for the two test versions, the data fall well within the commonly accepted ranges, which made the data suitable for further analysis. The Kolmogorov–Smirnov and Shapiro–Wilks tests of normality for the two test versions, however, indicated significantly non-normal distributions of data. Because the deviation of normality was not severe, the entire complement of data was used. The sample size of 881 for ECT version 1.3 might have improved the accuracy

of the Cronbach's alpha as the data were not normally distributed (Sheng & Sheng, 2012).

Graphical display of performance in items of the test

The scatter plot in Figure 1 displays the responses of the individuals who completed ECT version 1.2. It can be observed that most of the individuals answered the items correctly (60% of the responses to all the items were correct), while a smaller percentage provided incorrect responses (40% of the responses were incorrect). There was, however, a clearly significant increase in incorrect responses between items 19 and 24, and between items 36 and 39. This could be attributed to individuals choosing to answer certain items or not having sufficient time to correctly answer certain items in the test. The items' difficulty levels can only be confirmed by conducting an item analysis, however, and this was not done.

Figure 2 shows the responses in ECT version 1.3, and as the data on the answer sheet were scanned automatically, it meant that the missing responses were not captured. A similar trend was observed with respect to the correct responses, while the incorrect responses increased in items 23, 24, 26, 27, 39, 40, 41 and 42. The range of incorrect and correct responses for ECT 1.3 indicated that 62% of the responses to the items were correct, while 38% of the responses were incorrect. The pattern of incorrect and correct responses across the two test versions would suggest that perhaps the individuals completing the tests had intentionally skipped certain items in the test and did not necessarily need more time to complete the test.

Recorded test times for English comprehension test 1.3

The time that it took the last person in the different groups to complete ECT version 1.3 was recorded. Table 5 shows the different times recorded for the pilot run and the length of time it took the last person in each group to complete the ECT.

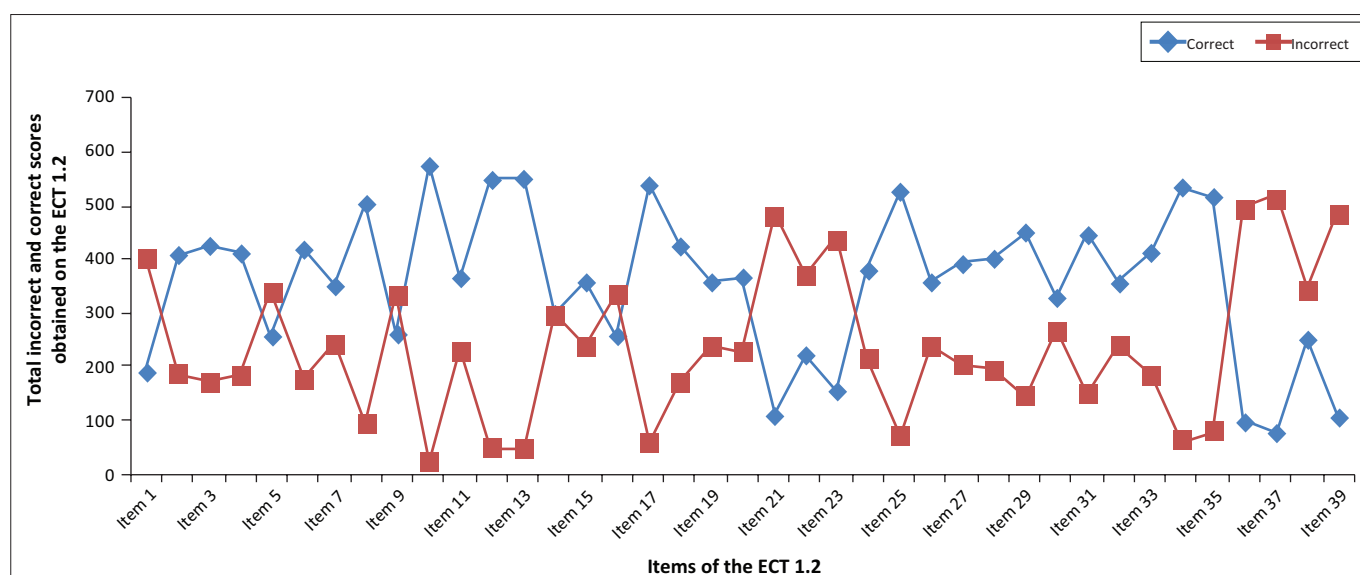


FIGURE 1: The performance of individuals in English comprehension test version 1.2.

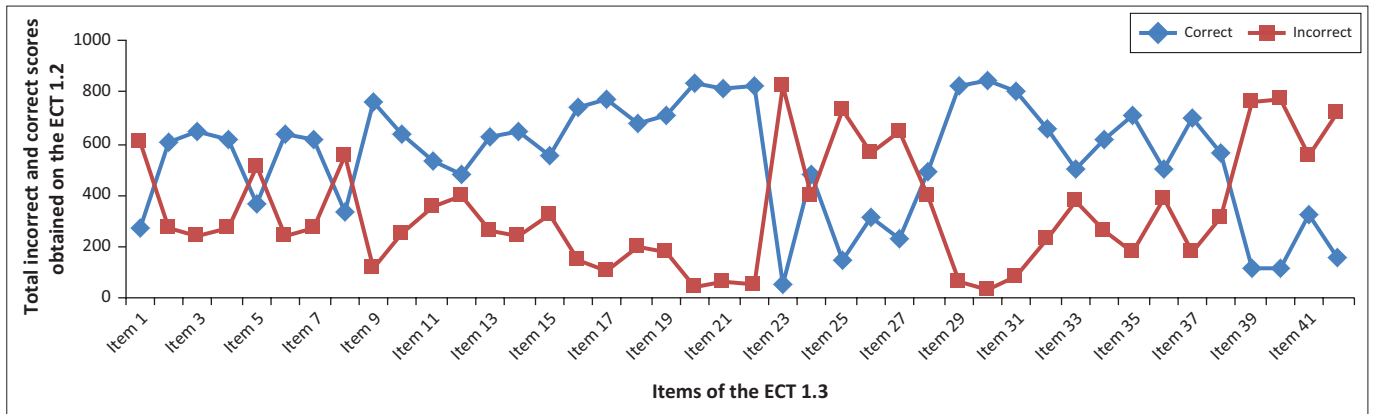


FIGURE 2: The performance of individuals on English comprehension test version 1.3.

TABLE 5: Test times for the pilot run of English comprehension test version 1.3.

Selections	Time
Pilot 1	79 min
Pilot 2	76 min
Pilot 3	70 min
Pilot 4	65 min
Pilot 5	80 min
Pilot 6	82 min
Pilot 7	79 min
Pilot 8	71 min
Pilot 9	76 min
Pilot 10	75 min
Pilot 11	71 min
Pilot 12	106 min
Pilot 13	70 min
Pilot 14	113 min
Pilot 15	61 min
Pilot 16	61 min
Pilot 17	59 min
Pilot 18	70 min
Pilot 19	69 min
Pilot 20	81 min
Pilot 21	73 min
Pilot 22	67 min
Pilot 23	84 min
Pilot 24	67 min
Pilot 25	68 min
Pilot 26	55 min
Pilot 27	75 min
Pilot 28	85 min
Pilot 29	63 min
Average time	74 min

From the times captured in Table 5, it is apparent that the candidates completed the test at different times in the 29 pilot tests that had been conducted, with an average of 74 min as completion time. The shortest time recorded was 55 min and the longest time recorded was 113 min. The fact that the last person in each group did not complete the test within 45 min is worth noting and it suggests that the set time limit of 45 min in ECT version 1.2 might be an unsuitable time limit.

Reliability results

The reliability coefficients for the two test versions are presented in Tables 6 and 9. The average scores as well as the

TABLE 6: Reliability statistics for English comprehension test 1.2.

Test version	Cronbach's alpha	Total test items	Mean
ECT version 1.2	0.789	39	23
Revised ECT version 1.2	0.820	30	17

Source: Arendse, D.E. (2018). *Exploring the construct validity and reliability of the English comprehension test*. (Unpublished doctoral thesis). Cape Town: University of Pretoria.
ECT, English comprehension test.

TABLE 7: Items removed from English comprehension test 1.2.

Removed items	Cronbach's alpha
Item 16	0.795
Item 17	0.799
Item 19	0.805
Item 18	0.811
Item 11	0.813
Item 15	0.816
Item 6	0.817
Item 10	0.817
Item 12	0.818

TABLE 8: Items lowering the Cronbach's alpha for English comprehension test version 1.2.

Item number	Item Description
6.	Which statement is FALSE according to the information given in the passage?
10.	(A. they, B. whose, C. who, D. which)
11.	(A. what, B. whom, C. whose, D. who)
12.	(A. is, B. were, C. was, D. has)
15.	WO Moses was thanked by Lt Col Hoekstra for helping him with the Aviation Safety course. (Fact/Opinion)
16.	WO Moses recognised the unsafe practices because he was a pilot and did the Aviation Safety course. (Fact/Opinion)
17.	There was a lot of rubbish and litter lying around on the airport strip. (Fact/Opinion)
18.	There were landmines in the area that the passengers had landed in. (Fact/Opinion)
19.	The passengers smoked when they landed because the flight was long and made them tired. (Fact/Opinion)

total items are indicated to place the reliability in Tables 6 and 9 in context. The reliability of the full test items is similar across the two test versions (see Tables 6 and 9), which may be regarded as acceptable reliability values for research purposes (Nunnally & Bernstein, 1994). To assess the best reliability coefficient for the data, the item total statistics were reviewed. These statistics highlighted the items that decreased the reliability coefficient value. The aforementioned items are

TABLE 9: Reliability statistics for English comprehension test 1.3.

Test version	Cronbach's alpha	Total test items	Mean
ECT version 1.3	0.785	42	26
Revised ECT version 1.3	0.816	33	20.94

Source: Arendse, D.E. (2018). *Exploring the construct validity and reliability of the English comprehension test*. (Unpublished doctoral thesis). Cape Town: University of Pretoria

ECT, English comprehension test.

TABLE 10: The items removed for English comprehension test 1.3.

Removed item	Cronbach's alpha
Item 8	0.789
Item 11	0.794
Item 9	0.798
Item 10	0.802
Item 6	0.806
Item 18	0.809
Item 23	0.811
Item 7	0.812
Item 25	0.813

TABLE 11: Items lowering the Cronbach's alpha for English comprehension test version 1.3.

Item number	Item Description
6.	Which statement is FALSE according to the information given in the passage?
7.	WO Moses was thanked by Lt Col Hoekstra for helping him with the Aviation Safety course. (Fact/Opinion)
8.	WO Moses recognised the unsafe practices because he was a pilot and did the Aviation Safety course. (Fact/Opinion)
9.	There was a lot of rubbish and litter lying around on the airport strip. (Fact/Opinion)
10.	There were landmines in the area that the passengers had landed in. (Fact/Opinion)
11.	The passengers smoked when they landed because the flight was long and made them tired. (Fact/Opinion)
18.	(A. which, B. whom, C. who, D. whose)
23.	(A. was, B. were, C. is, D. has been)
25.	Solicit (Synonym)

indicated in Tables 7 and 8 for ECT 1.2 and in Tables 10 and 11 for ECT 1.3. For the best coefficient to be obtained, the items that decreased the reliability coefficient were deleted and the reliability analysis was rerun. This process was repeated until the reliability coefficient was at its highest value, which is depicted in Table 6 for ECT 1.2 and Table 9 for ECT 1.3.

In ECT version 1.2 (Tables 6 and 7), items 6, 10, 11, 12, 15, 16, 17, 18 and 19 were deleted to improve the reliability coefficient. The reliability coefficient on standardised items in the remaining 30 items was 0.820, indicating an acceptable reliability (Nunnally & Bernstein, 1994). This is, however, still insufficiently reliable for selection purposes or high-stakes testing (Foxcroft & Roodt, 2009). The mean of these 30 items was 17, which suggests that, on average, individuals answered 44% of the test correctly.

In Table 8, the contents of the items lowering the Cronbach's alpha are indicated. The varied contents of the items indicated in Table 8 allow one to infer that these items were possibly affected by both ability and speed.

For ECT version 1.3 (Tables 9 and 10), items 6, 7, 8, 9, 10, 11, 18, 23 and 25 were deleted to improve the reliability statistic.

The 33 remaining items produced a reliability statistic of 0.816 on standardised items, indicating an acceptable reliability (Nunnally & Bernstein, 1994). It is, however, inadequate for selection purposes or high-stakes testing (Foxcroft & Roodt, 2009). The mean of these 33 items was 21, which indicates that, on average, individuals correctly answered 50% of the test questions.

In Table 11, the contents of the items lowering the Cronbach's alpha are indicated. The contents of the items are varied in Table 11 and were possibly affected by the ability as speed was not a factor affecting ECT version 1.3.

The items of the ECT 1.2 and ECT 1.3 (Tables 7, 8, 10 and 11) that lowered the Cronbach's alpha were negatively affecting the intercorrelations of the test and lowering the internal consistency of the test (Streiner, 2003). Moreover, the deletion of items that lowered the Cronbach's alpha was necessary as these items were malfunctioning despite the fact that individuals had a longer time within which to complete them. This raises the dilemma between speed and ability for the ECT 1.2, while ECT 1.3 could have been predominantly affected by ability.

Discussion

The biographical details of the sample were taken into consideration as they informed the context of the results. The sample was dominated by men, particularly under the age of 25 years, who spoke an African language. This suggests that women and all language groups were not equally represented, which is a limitation in convenience sampling. The implication of this specific sample is that the overwhelming majority were non-native English speakers. This is crucial when considering that the ECT is in English and thus language is an inherent variable that could contribute to measurement error in the calculation of the Cronbach's alpha of the two ECT versions (Dingwall et al., 2014, 2017; Kanniainen et al., 2019; Nel, 2018; Spaul, 2016; Van de Vijver & Rothmann, 2004).

The substantial differences in the test administration, test structure and instructions of the two test versions may also have had an impact on the reliability of the ECT. The minimum scores of individuals in the ECT correlate with the reading comprehension and literacy concerns raised by researchers (Dingwall et al., 2014, 2017; Kanniainen et al., 2019; Nel, 2018; Spaul, 2013). These minimum scores could also be influenced by the manner in which items were phrased or the level of complexity of the items (Dingwall, et al., 2014, 2017). When comparing the minimum and maximum scores of the timed (8 and 38) and untimed (8 and 39) versions of the ECT, it would appear that these scores were not adversely affected by the time limit (Angelidis et al., 2019; Bunyi et al., 2015; Keith & Reynolds, 2010; Oberauer & Lewandowsky, 2013). Although time limits are occasionally required to assess ability, the absence of a time limit may sometimes overestimate ability (Keith & Reynolds, 2010; Oberauer & Lewandowsky, 2013). The time limit imposed for ECT version 1.2 may affect reliability because of the compromise between speed and

ability (Goldhammer, 2015). Because the intention of the ECT is to act as a screening tool, it does not require a time limit as the aim was to establish a baseline of ability, specifically verbal reasoning. Although factors such as literacy and reading comprehension are worth considering when measuring verbal reasoning, there are other factors such as working memory, coding or decoding and reasoning skills that are equally important to consider (Asgari & Schutze, 2017; Keith & Reynolds, 2010; Lohman & Lakin, 2009; Oberauer & Lewandowsky, 2013). Although individuals require literacy skills when reading and comprehending texts, they also use their working memory, coding or decoding processes and reasoning skills to reach valid conclusions (Asgari & Schutze, 2017; Lohman & Lakin, 2009). These cognitive processes and skills can be impacted by speed and may affect the reliability of the ECT. The context in which the ECT can be used, either educational or organisational, does not necessarily require timed screening. For this reason, the measurement of ability supersedes the use of speed.

The recorded times were based on the time that the last person completed the test, with the average time taken being 74 min, which was 29 min longer than the time limit of 45 min that was imposed for ECT version 1.2. The removal of the time limit and recording the time the last person finished can be regarded as a form of accommodation of participants to support the extraction of ability for the slowest persons (Kuwornu, 2017). Owing to the awareness that the ECT sample comprised predominantly non-native English speakers, mechanisms such as accommodation were required as the time limit might have placed the focus on the items completed instead of the measurement of the construct (Keith & Reynolds, 2010; Oberauer & Lewandowsky, 2013).

The observation of incorrect and correct responses throughout the two test versions would suggest that individuals preferred to answer certain items and were therefore less affected by the time limit. This therefore emphasises the compromise between speed and ability (Goldhammer, 2015). In the administration of ECT version 1.3, it was qualitatively observed that most candidates completing the test would spend the majority of their time on the last section of the test. The last section of the test contained sentence construction items and these items could therefore have been the reason for the long time spent on the test. There is, however, no quantifiable evidence to support this qualitative observation that was noted during the testing.

The Cronbach's alpha for the full item scale of both test versions was appropriate for research purposes but insufficient for high-stakes selection purposes (Nunnally & Bernstein, 1994). When some items that reduced the Cronbach's alpha values were removed, the Cronbach's alpha for the revised test versions was sufficient for measuring ability across the two test versions (Nunnally & Bernstein, 1994).

From the examination of the items that lowered reliability across the two test versions (Tables 8 and 11), it is clear that

there are some identical items. The identical items across the two test versions are the following: one 'False' and five 'Opinion and Fact' items. These items either lowered reliability because the items in relation to the comprehension section to which it refers may not be clear or the distractors for these items created inconsistency in answering patterns. Because these items are based on comprehension and the comprehension section was not removed from either test version, one might ponder whether the outcome could be because of poor reading comprehension skills and literacy (Abedi, 2002; Bahardoost & Ahmadi, 2018; Dingwall et al., 2014, 2017; Howie et al., 2017; Kanninen et al., 2019; Nel, 2018; Spaul, 2016; Streiner, 2003). Moreover, comprehension-based items may affect reliability as the items are dependent on the individual understanding the comprehension piece (Streiner, 2003). These items in the ECT that are based on the comprehension piece are not dependent on each other; however, each item assesses different inferences regarding the comprehension piece. It is nevertheless worth considering that the participants' understanding of the comprehension had a direct impact on their ability to respond to items that depend on inferences in the comprehension (Bahardoost & Ahmadi, 2018; Dingwall et al., 2014, 2017; Kanninen et al., 2019; Nel, 2018; Streiner, 2003). Because the majority of the sample were non-native English speakers, the content of the test and items could have been more challenging in terms of the language used and questions posed (Abedi, 2002; Dingwall et al., 2014, 2017; Van de Vijver & Rothmann, 2004). The responses to the comprehension-based items could also have been affected by external factors such as SES and the quality of education received (Cockcroft et al., 2016; Spaul, 2016). These external factors may also include personal contexts, urban and rural living circumstances, family history and traditional understanding, which may have influenced how individuals responded to these comprehension items.

The remaining items were different 'tense' items across the two test versions that were identified as lowering the Cronbach's alpha (see Tables 8 and 11). These 'tense' items were not related to the comprehension piece as they were separate, grammar-related, language questions. These 'tense' items could therefore either be too challenging as they required more formal English knowledge that African- and Afrikaans-language individuals might not have, depending on their school background (Abedi, 2002; Cockcroft et al., 2016; Dingwall et al., 2014, 2017; Krugal & Fourie, 2014; Kuwornu, 2017; Pretorius & Klapwijk, 2016; Spaul, 2016; Van de Vijver & Rothmann, 2004). The handling of 'tense' could also be an issue affected by low literacy levels, differing uses of tense across languages, errors in forward or back translation processes and decoding errors on specific tense terms (Asgari & Schutze, 2017; Bahardoost & Ahmadi, 2018; Dingwall et al., 2014, 2017; Howie et al., 2017; Kanninen et al., 2019; Nel, 2018; Spaul, 2016; Streiner, 2003).

One synonym item that was identified in ECT version 1.3 was also a language item but it did not relate to the comprehension piece. This item could be affected by participants' vocabulary

and literacy knowledge (Abedi, 2002; Cockcroft et al., 2016; Krugal & Fourie, 2014; Kuwornu, 2017; Pretorius & Klapwijk, 2016). Moreover, language generally affects the performance of non-native English speakers in English assessments (Abedi, 2002; Dingwall et al., 2014, 2017; Kuwornu, 2017; Van de Vijver & Rothmann, 2004). However, it is recommended that the content of these items should be explored in greater depth and in accordance with the principles of linguistic literature to establish language-related issues that may affect non-native English speakers.

The reliability of the ECT was nevertheless not negatively influenced by either the timed or untimed versions of the ECT. Moreover, the internal consistency of the two test versions appears to be acceptable, particularly the revised test versions. This acceptable internal consistency indicates that most of the items across the test versions appear to measure the same construct consistently. Cognisant of this, the current study suggests that the internal consistency of the ECT across the test versions was not negatively affected by time but this does not mean that performance in the two test versions was not affected. The removal of items that lowered the Cronbach's alpha was necessary if one considered that these items were possibly not related to the construct being measured, and were affecting the unidimensionality of the test. Furthermore, these items have lower inter-relations with other items in the test and thus lowered the internal consistency of the test (Streiner, 2003). It is, however, possible that Cronbach's alpha was underestimated in both the initial and revised reliability analysis, as the true reliability could be much higher (Abedi, 2002; Osburn, 2000; Streiner, 2003; Taber, 2018; Tavakol & Dennick, 2011). This argument would suggest that the two test versions appear to be sufficiently reliable for research (Nunnally & Bernstein, 1994) and, when revised, it may be able to measure verbal reasoning consistently (Nunnally & Bernstein, 1994).

The performance of candidates across the two test versions was not assessed and may provide valuable insights into the ECT in future research. The two test versions had different numbers of items, which is an important consideration in the light of the reliability results. It should nevertheless be cautioned that although the results for the two test versions were consistent, this does not imply that the performance across the test versions was equal. It is crucial to obtain these results for further developing and refining the ECT. It thus opens up more avenues for research relating to the ECT.

There are a few important limitations concerning this study that should be noted. The samples for both test versions were conveniently selected. Therefore, the results cannot be generalised and are specific to the population that was utilised. The lack of missing scores in the analysis of incorrect and correct items on the scatter plot is a limitation in assessing the accurate number of incorrect items across the test versions, and thus the incorrect data are regarded as possibly being inflated. Another limitation of this study was that an average time for completing the untimed test version (ECT 1.3) could not be calculated because alternative times, such as

the time when the first person completed the test, were not recorded. The external factors such as stress, fatigue, motivation, anxiety, attitude and energy levels of participants, and internal test factors such as systematic errors, may have affected the reliability of both versions of the ECT.

It is recommended that in future piloting of the ECT the time should be recorded for the first and last persons to complete the ECT in order to establish a more accurate range of the time taken by individuals to complete the test. The recording of the first and last persons completing the test would allow for an average time to be calculated, which is a more accurate calculation of the time needed to complete the test. It is also recommended that the performance in the two ECT versions should be assessed to establish whether there was a difference in performance. Moreover, the performance of the nine African and Afrikaans language individuals who are non-native English speakers should be compared to English first-language speakers across test versions. Another recommendation is that the items identified as lowering the Cronbach's alpha should be explored in more detail in terms of the appropriate linguistic literature and statistical analysis. This may inform whether English, the nine African languages or Afrikaans language individuals perform differently in such items and find a possible reason why they would perform differently or similarly for these items.

Conclusion

This study embarked on assessing the reliability of individuals in the timed version (ECT 1.2) and the untimed version (ECT 1.3). The administration differences (including test structure and instructions) could have affected the reliability of the ECT. The recorded times indicated that the last person to complete the test was unable to complete it within 45 min, which was the time limit of the timed test version (ECT 1.2). The performance of individuals in the untimed and timed versions of the ECT appears to be similar according to the average minimum and maximum scores. This performance could be attributed to the answering pattern of individuals, when they might deliberately have chosen to answer certain items and therefore might not have needed more time for answering test items. This more importantly suggests the unsuitability of a time limit for the ECT, as the compromise between speed and ability affects the reliability of the test. The reliability results indicated that both tests were appropriate for research purposes and once the items that lowered the Cronbach's alpha had been removed, both test versions were able to measure the verbal reasoning aspect of the ECT consistently. The revised reliability results across the two test versions suggested that the internal consistency was acceptable. Removal of the items lowering the Cronbach's alpha across test versions was important as they negatively affected the reliability and internal consistency of the test. This study provides important information on the psychometric properties of the ECT and is imperative for further development of the ECT.

Acknowledgements

The author acknowledges that this article is related to some of the findings from her thesis published in 2018, which is entitled: 'Exploring the construct validity and reliability of the English comprehension test' (University of Pretoria).

Competing interests

The author has no competing interests. She also declares that she has no financial or personal relationship that may have inappropriately influenced her in writing this article.

Authors' contributions

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

Funding information

There was no funding received for the publishing of this article.

Data availability statement

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Disclaimer

The views and opinions expressed in this article are the author's own and are not the official position of any institution.

References

- Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometrics issues. *Educational Assessment*, 8(3), 231–257. https://doi.org/10.1207/S15326977EA0803_02
- Angelidis, A., Solis, E., Lautenbach, F., Van der Does, W., & Putman, P. (2019). I'm going to fail! Acute cognitive performance anxiety increases threat-interference and impairs WM performance. *PLoS One*, 14(2), 1–32. <https://doi.org/10.1371/journal.pone.0210824>
- Arendse, D.E. (2018). *Exploring the construct validity and reliability of the English comprehension test* (Unpublished doctoral thesis). Pretoria: University of Pretoria.
- Arendse, D.E., & Maree, D. (2019). Exploring the factor structure of the English comprehension test. *South African Journal of Psychology*, 49(3), 376–390. <https://doi.org/10.1177/0081246318805268>
- Asgari, E., & Schutze, H. (2017, September 7–11). Past, present, future: A computational investigation of the typology of tense in 1000 languages. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 113–124). Copenhagen: Association for Computational Linguistics.
- Bahardoost, M., & Ahmadi, A. (2018). The relationship between test-taking strategies and Iranian EFL learners' performance on reading comprehension tests. *International Journal of Foreign Language Teaching and Research*, 6(22), 117–130.
- Bekwa, N.N. (2016). *The development and evaluation of Africanised items for multicultural cognitive assessment* (Unpublished doctoral thesis). University of South Africa, Pretoria.
- Bunyi, J., Heal, C., Nadendla, S., Sherman, D., Tran, J.D., & Varsos, J. (2015). Test scores on timed exams decline over time without a significant increase in physiological stress. Timed test performance and physiological stress. *Journal of Advanced Student Science*, 2015 (1, Spring), 1–21.
- Cockcroft, K., Bloch, L., & Moolla, A. (2016). Assessing verbal functioning in South African school beginners from diverse socioeconomic backgrounds: A comparison between verbal working memory and vocabulary measures. *Education as Change*, 20(1), 199–215. <https://doi.org/10.17159/1947-9417/2016/559>
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334. <https://doi.org/10.1007/BF02310555>
- Dingwall, K.M., Gray, A.O., McCarthy, A.R., Delima, J.F., & Bowden, S.C. (2017). Exploring the reliability and acceptability of cognitive tests for indigenous Australians: A pilot study. *BioMed Central (BMC) Psychology*, 5(26), 1–16. <https://doi.org/10.1186/s40359-017-0195-y>
- Dingwall, K.M., Lindeman, M.A., & Cairney, S. (2014). 'You've got to make it relevant': Barriers and ways forward for assessing cognition in Aboriginal clients. *BioMed Central (BMC) Psychology*, 2(13), 1–11. <https://doi.org/10.1186/2050-7283-2-13>
- Dodeen, H.M., Abdelfattah, F., & Alshumrani, S. (2014). Test-taking skills of secondary students: The relationship with motivation, attitudes, anxiety and attitudes towards tests. *South African Journal of Education*, 34(2), 1–18. <https://doi.org/10.15700/201412071153>
- Erguven, M. (2014). Two approaches to psychometric process: Classical test theory and item response theory. *Journal of Education*, 2(2), 23–30.
- Field, A.P. (2009). *Discovering statistics using SPSS*. London: Sage.
- Foxcroft, C.D., & Roodt, G. (2009). *An introduction to psychological assessment in South African context*. Cape Town: Oxford University Press.
- Goldhammer, F. (2015). Measuring ability, speed, or both? Challenges, psychometric solutions, and what can be gained from experimental control. *Measurement*, 13(3–4), 133–164. <https://doi.org/10.1080/15366367.2015.1100020>
- Guttman, L. (1945). A basis for analysing test-retest reliability. *Psychometrika*, 10(4), 255–282.
- Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behaviour Research*, 50, 1166–1186. <https://doi.org/10.3758/s13428-017-0935-1>
- Howie, S.J., Combrink, C., Roux, K., Tshela, M., Mokoena, G.M., & Mcleod Palane, N. (2017). *PIRLS literacy 2016: South African highlights report*. Pretoria: Centre for Evaluation & Assessment.
- Kanniainen, L., Kili, C., Tolvanen, A., Aro, M., & Leppanen, P.H.T. (2019). Literacy skills and online research and comprehension: Struggling readers face difficulties online. *Reading and Writing*, 32(9), 2201–2222. <https://doi.org/10.1007/s11145-019-0944-9>
- Keith, T.Z., & Reynolds, M.R. (2010). Cattell-Horn-Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools*, 47(7), 635–650. <https://doi.org/10.1002/pits.20496>
- Kendeou, P., Van den Broek, P., Helder, A., & Karlsson, J. (2014). A cognitive view of reading comprehension: Implications for reading difficulties. *Learning Disabilities Research and Practise*, 29(1), 10–16. <https://doi.org/10.1111/ldrp.12025>
- Kiwan, D., Ahmed, A., & Pollitt, A. (2000). *The effects of time-induced stress on making inferences in text comprehension*. Paper presented at the European Conference on Educational Research, Edinburgh.
- Krugal, R., & Fourie, E. (2014). Concerns for the language skills of South African learners and their teachers. *International Journal of Educational Science*, 7(1), 219–228. <https://doi.org/10.1080/09751122.2014.11890184>
- Kuwornu, A.A. (2017). Review of issues of language assessments for non-native speakers of English. *Sino-US English Teaching*, 14(3), 157–168. <https://doi.org/10.17265/1539-8072/2017.03.005>
- Lee, Y., & Chen, H. (2011). A review of recent response-time analyses in educational testing. *Psychological Test and Assessment Modelling*, 53(3), 359–379.
- Liao, Y. (2004). Issues of validity and reliability in second language performance assessment. Teachers College, Columbia University. *Working Papers in TESOL and Applied Linguistics*, 4(2), 1–4.
- Lohman, D.F., & Lakin, J.M. (2009). Reasoning and intelligence. In R.J. Sternberg & S.B. Kaufman (Eds.), *Handbook of intelligence* (2nd edn., pp. 1–47). New York: Cambridge University Press.
- Mushquash, C.J., & Bova, D.L. (2007). Cross-cultural assessment and measurement issues. *Journal of Developmental Disabilities*, 13(1), 55–66.
- Nel, C. (2018). A blueprint for data-based English reading literacy instructional decision-making. *South African Journal of Childhood Education*, 8(1), 1–9. <https://doi.org/10.4102/sajce.v8i1.528>
- Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory* (3rd edn.). New York: McGraw-Hill.
- Oberauer, K., & Lewandowsky, S. (2013). Evidence against decay in verbal working memory. *Journal of Experimental Psychology*, 142(2), 380–411. <https://doi.org/10.1037/a0029588>
- Osborn, H.G. (2000). Coefficient alpha and related internal consistency reliability coefficient. *Psychological Methods*, 5(3), 343–355. <https://doi.org/10.1037/1082-989X.5.3.343>
- Pretorius, E.J. (2002). Reading ability and academic performance in South Africa: Are we fiddling while Rome is burning? *Language Matters*, 33, 169–196. <https://doi.org/10.1080/10228190208566183>
- Pretorius, E.J., & Klapwijk, N.M. (2016). Reading comprehension in South African schools: Are teachers getting it, and getting it right? *Per Linguam*, 32(1), 1–20.
- Santos, J.R.A. (1999). Cronbach's alpha: A tool for assessing the reliability of scales. *Journal of Extension*, 37(2), 1–6.
- Sheng, Y., & Sheng, Z. (2012). Is coefficient alpha robust to non-normal data? *Frontiers in Psychology*, 3, 1–13. <https://doi.org/10.3389/fpsyg.2012.00034>
- Spaull, N. (2013). *South Africa's education crisis: The quality of education in South Africa 1994–2011*. Johannesburg: Centre for Development and Enterprise.
- Spaull, N. (2016). *What do we know about reading outcomes in South Africa?* Paper presented at the Bridge Forum, Johannesburg.
- Streiner, D.L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80(1), 99–103. https://doi.org/10.1207/S15327752JPA8001_18
- Taber, K.S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research Science Education*, 48 (6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. <https://doi.org/10.5116/ijme.4dfb.8dfd>
- Van de Vijver, F.J.R., & Rothmann, S. (2004). Assessment in multicultural groups. The South African case. *South African Journal of Industrial Psychology*, 30(4), 1–7. <https://doi.org/10.4102/sajip.v30i4.169>