



Jones, M. J., & Smith, M. (2014). Traditional and alternative methods of measuring the understandability of accounting narratives. *Accounting, Auditing and Accountability Journal* , 27 (1), 183-208. DOI: 10.1108/AAAJ-04-2013-1314

Peer reviewed version

Link to published version (if available):  
[10.1108/AAAJ-04-2013-1314](https://doi.org/10.1108/AAAJ-04-2013-1314)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Emerald at <http://www.emeraldinsight.com/doi/abs/10.1108/AAAJ-04-2013-1314>. Please refer to any applicable terms of use of the publisher.

## **University of Bristol - Explore Bristol Research**

### **General rights**

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/pure/about/ebr-terms.html>

# Investigating new methods of measuring the understandability of accounting narratives.

Michael John. Jones\* and Malcolm Smith\*\*

\* University of Bristol

\*\*Edith Cowan University, Western Australia

## **Acknowledgements**

We recognise the research assistance of Jacqui Whale in the construction of this paper. We would also like to thank participants at the 14<sup>th</sup> Annual Financial Reporting and Business Communication conference in Bristol in July 2010, particularly the discussant Garcia Beatriz Osman. We would also like to thank participants at seminars given at the British Accounting and Finance Conference in April 2011 in Birmingham, UK and at the European Accounting Association in Rome in April 2011. Finally, we would like to thank the two reviewers of this paper for their informed comments.

## ABSTRACT

**Purpose** The study explores the use of alternative measures of ‘understandability’ on accounting texts. This includes the Meaning Identification test (MIT) and the Sentence Verification Technique (SVT), which have not previously been used by accounting researchers, as well as variants on the traditional CLOZE tests such as the C-Test.

**Design/methodology/approach** The prior literature in accounting has strongly argued that readability and understandability are different, if related topics. There has also been considerable criticism of the existing measures, such as the Flesch test and Cloze Procedure, which have been used to assess the readability and understandability of accounting narratives. This study focuses on the understandability of accounting texts and evaluates how reliable the Cloze test is as a measure of comprehension. An experiment was designed and conducted to measure users’ performance with a variety of comprehension measures.

**Findings** The study concludes that the outcomes from the MIT and SVT comprehension tests are not significantly associated with those from the CLOZE tests. This implies that while the Cloze test is a good measure of the predictability of accounting narratives, and of textual redundancy, it does not necessarily measure the understandability of the text.

**Originality/value** These measures of understandability, which have not previously been used in accounting, would enable researchers to test the communicational effectiveness of using different accounting narratives such as annual reports or prospectuses. Such a measure could be used to improve the understandability of accounting narratives. The strengths and weaknesses of the various tests are assessed. It is suggested that there is a need for further experimentation especially with the MIT test.

**Keywords** comprehension, measurement, understandability

**Paper type** research paper

# Investigating new methods of measuring the understandability of accounting narratives.

## 1. Introduction

Accounting narratives have grown in importance over the past twenty years (e.g., Bryan, 1997; Arthur Andersen, 2001; Aerts, 2010). They represent opportunities for management to describe, discuss and evaluate the financial and non-financial performance of the company. Accounting narratives commonly found in annual reports are the Chairman's statement, Chief Executive's Review and Business Review. These narratives set the context for the financial statements.

An important aspect of these narratives is the need for effective communication in which there is a shared meaning between the producers and users of the information (Smith and Taffler, 1992a). Effective communication occurs in financial reporting only if the meanings intended by the source of the information are correctly assigned to the financial statement messages by the destination (Smith and Smith, 1971). Effective communication is essential if financial information, such as the annual report, is to be understood. Without such an understanding, financial information will not be processed efficiently and effectively and prove a relevant and reliable basis for either stewardship or financial decision-making. There have been a wealth of studies into the effective communication of financial information in the annual reports (e.g. Parker, 1982; Smith & Taffler, 2000; Clatworthy & Jones, 2001) with current concerns relating to issues of obfuscation and impression management (e.g., respectively Courtis, 1998; Merkl-Davies & Brennan, 2007) and their potential to impact negatively on effective communication.

In particular, there has been a focus on two related aspects of effective communication: readability and understandability. The early literature made no distinction between the two, assuming that readability and understandability measured the same communication attributes (Adelberg, 1983; Adelberg and Razek, 1984). More recent studies (e.g. Smith and Taffler, 1992b; Jones and Shoemaker, 1994), however, distinguish between the two concepts. Jones (1997, p. 105) argues: "Readability thus measures the textual difficulty of a passage; while understandability measures the ability of a reader to gain knowledge from a text, and is contingent not only on syntactical difficulty, but also on reader characteristics such as the

reader's background, prior knowledge, interest and general reading ability". While readability can be measured satisfactorily in terms of word length, sentence length and complexity of content (e.g., through Flesch, FOG and Lix indices), Smith and Taffler (1992a) suggest that it is dangerous to use such measures as proxies for 'understandability'. Courtis (2004) suggests that syntactical complexity is a proxy for obfuscation.

Reader-text interactive texts, principally the Cloze procedure have been used to gauge understandability. Essentially, readability is passive and text-centred, while understandability is interactive and reader centred. Kintch and Vipond (1979) have developed alternative formulas consistent with psycholinguistic factors of reading comprehension, but these serve to demonstrate how well traditional formulas work despite their weak construct validity. Brennan, Guillamon-Saorin and Pierce (2009) note that many authors have acknowledged the limitations of the application of readability formulas to accounting narratives. Courtis (1996) notes the potential importance of repetition of information in the understandability of financial reports.

However, despite 40 years of attempting to measure the readability and understandability of accounting narratives, there is no agreed measure of understandability. This paper seeks to throw new light on the topic of understandability; in particular, it aims to help us find a stronger and more robust measure by which we can attempt to assess the textual difficulty of accounting narratives such as typically found in the annual report. This will help us better to gauge whether or not the accounting narratives are effective means of communicating financial information. If they are not a new measure could help improve their understandability. We test three new measures of textual comprehension, (the C-Test, the MIT test and SVT test) and we evaluate them against an existing measure, the Cloze test. We use an experimental setting and administer the test to 44 UK final year undergraduate students under exam conditions. In particular, we investigate the strengths and weaknesses of these measures as potential ways of measuring the understandability of accounting narratives.

The rest of the paper is structured in four parts followed by a conclusion. In the next section, we review the prior literature in assessing the communicational effectiveness of accounting narratives distinguishing between the 'readability' and 'understandability' studies. Then, in section three, we outline three new tests that we explore in this paper: the C-Test, the Sentence Verification technique and the Meaning Identification Technique. Section four

outlines the methods of our study and sets out how we constructed and administered the tests. Then, in section five, we discuss our findings in terms of the results and the ease of development and administration of the tests.

## **2. Prior Accounting Literature into Readability and Understandability.**

Accounting narrative research has a long pedigree stretching back to the 1970s. It has been categorised into two main streams in a comprehensive overview of the then extant literature by Jones and Shoemaker (1994): content analytic studies and syntactical studies. This broad dichotomy still holds today with content studies investigating the thematic structures and patterns within the narratives and the syntactical studies investigating the interrelated, but distinct, topics of readability and understandability. It is the syntactical studies which are the focus of this particular study.

The early syntactical studies made no real distinction between the readability and the understandability of accounting narratives. Indeed, early researchers such as Taylor (1953) or Klare (1963) used the terms ‘readable’ and ‘understandable’ as broadly synonymous. However, later studies such as Smith and Taffler (1992b), Jones and Shoemaker (1994, p. 172) and Clatworthy and Jones (2001) concur stating that there are two main approaches to measuring textual difficulty: the first using syntactical difficulty and the second using “sophisticated psycholinguistic and socio-linguistic techniques such as Cloze, multidimensional scaling, association analysis and classification analysis”. The former are usually termed “readability studies” while the latter are normally termed understandability studies.

This research study follows the prior literature and its focus is on measuring understandability not readability. To date, the focus of recent interest in accounting has generally focused on the Cloze procedure as a method of measuring understandability. The Cloze procedure, in effect, uses the principle that if individuals can understand a piece of text they will be able to fill in missing words correctly. Thus, in standard Cloze procedure tests, every fifth or tenth word in a narrative is deleted. The Cloze procedure is a form of comprehension test which determines how well the essential information in the document comes across to the reader (de Jong & Schellens, 1997) and has been used by accounting researchers as a proxy for understandability. Jones (1997) investigated nine empirical studies

that had used the Cloze Procedure to assess the effectiveness of accounting communications such as accounting and tax textbooks (e.g. Adelberg and Razek, 1984; Raabe, Stevens and Stevens, 1984), authoritative pronouncements (e.g. Shaffer, Stevens and Stevens, 1983) and corporate annual reports (Smith and Taffler, 1992a, 1992b). Smith and Taffler's studies are particularly pertinent to this study and their findings may be taken as illustrative. Smith and Taffler, (1992a, 1992b) used both accounting undergraduates and then accounting practitioners to assess extracts from the chairman's narratives of failed and non-failed companies. Smith and Taffler (1992a, 1992b) had two major findings. First, they determined that accounting practitioners and accounting undergraduates did not understand ten out of 18 and 15 out of 18 passages, respectively. Second, they found a poor relationship between the Cloze scores and text-based scores such as Flesch and Lix. They thus conclude that the Cloze procedure measures understandability while Flesch measures readability (Smith and Taffler, 1992b, p.93).

However, there is considerable doubt as to whether the Cloze Procedure does indeed measure understandability. The Cloze procedure was developed by Taylor (1953) as "a new psychological tool for measuring the effectiveness of communication" (1953, p. 415). The Cloze procedure relies upon the human tendency to predict text from what has gone before taking contextual factors into account. For example, in the phrase "the cat sat on the ---", most readers would read "mat" for the deletion" (Jones, 1997, p. 107). This perhaps suggests that the Cloze procedure may measure textual redundancy which may or may not equate with understandability.

Taylor (1953) argued that by deleting words systematically from different passages and asking subjects to guess from the context missing words that passages could be ranked according to readability. The greater the number of correctly guessed replacements the greater the 'readability' of the passage. Most pedagogic narratives have deleted every fifth word using 250-word passages (see Taylor 1953; Miller and Coleman, 1967). Various methodological variants have been tried with exact replacement, synonym scoring and dash-line Cloze (as many dashes as letters). Dash-line Cloze scored the highest with replacement and synonym scoring the same. McNamara, Kintsch, Butler-Songer and Kintsch (1996) argue that the high intercorrelation between readability and Cloze scores is inevitable, and a consequence of word excision on sentence length. Kintsch and Vipond (1979) question the

accuracy of the Cloze procedure as an indicator of understanding, other than an an indicator of intrasentential redundancy.

A particular methodological problem with the Cloze procedure (and indeed with the readability studies as well) is the exact meaning of the Cloze scores. In particular, much effort has been expended to find a criterion reference score with which Cloze scores can be compared (see, for example, Bormuth, 1967, 1968; Entin and Klare, 1978; Rankin and Culhane, 1969). In general, the Cloze scores have been benchmarked against results from multiple-choice scores: 75% (instructional level) and 90% (independent). The instructional level represents passages sufficiently understandable to subjects in supervised instruction; while independent level passages represent text sufficiently understandable so that subjects can answer 90% (Bormuth, 1968). Passages with less than 75% for instructional level material and less than 90% for independent level material are deemed not understandable. In accounting, annual reports would be independent level material, while textbooks, for example, would be instructional level material. The accounting research literature has relied upon Bormuth (1968)'s study which established benchmark criteria representing scores of 44% for instructional study and 57% for independent study as equivalent to a 75% score on multiple choice methods and for instructional study as equivalent to a 90% score. Thus, Cloze scores for accounting materials below 44% for instructional materials like textbooks are deemed not understandable, while for independent level materials, a score below 57% is deemed not understandable. Newton et al, (1998), for example, related the scores in their study as  $\leq 43\%$  frustration level, 44-56% instructional level and  $\geq 57\%$  independent level. They also indicate that easy-to-read texts should have at most 10% of readers scoring below the independent level.

The test itself is much more time consuming to construct and score than the readability tests, but fairly easy and economical to construct when compared to multiple choice questionnaires. However, it has been criticised on a number of grounds. First, there is still no clear consensus on what the Cloze procedure actually measures, and whether or not, it actually measures understandability. Royer (2004, p.26) suggests that most Cloze tests measure sentence comprehension rather than the comprehension of a narrative passage.

Abraham and Chapelle (1992)'s comment still holds: "The practice of using Cloze tests without clear knowledge of what they measure is inconsistent with current objectives of



educational measurement”. Second, Shanahan, Kamil and Tobin, (1982); Kintsch and Miller, (1984) suggest that the Cloze procedure measures the ability to guess correctly missing words rather than the comprehension per se of the passages. Third, the benchmark criteria points by which the Cloze procedure is measured vary widely in tests against multiple comprehension tests. Jones (1997, p.120) looked at five tests (Bormuth, 1967, 1968; Rankin and Culhane, 1969; Rankin 1971 and Entin and Klare 1978). Jones (1997, p. 119) concludes: “If their explanations are accepted, the appropriate range is between 36 percent and 44 percent for the instructional levels and between 50 per cent and 61 per cent for the independent levels. However, even with criterion reference bands, both Cohen (1975) and Bailey and Harrison, (1984) advise caution, particularly where specific skills or prior knowledge are necessary to understand the context”. Fourth, doubts have been expressed as to its validity as a testing device (Klein-Braley and Raatz, 1984; Weir, 1990), and also how well it translates to a measure of a document’s “functional communication” (de Jong and Schellens, 1997). Finally, the test has also been criticized operationally as being an irritating (Weir 1990) or frustrating (Klein-Braley and Raatz) test to take.

The prior work into the readability and understandability literature thus demonstrates severe limitations with the current methodological approaches. This particular study examines several alternative measures of comprehension and assesses the Cloze procedure against them. In particular, it uses the C-test, the Sentence Verification Technique (SVT) and the Meaning Identification Technique (MIT). All of these are comparatively new techniques when compared with the Flesch test and the Cloze procedure although they are still over 30 years old. Perhaps, surprisingly, they have not, to the authors’ knowledge, so far been used in the accounting domain.

### **3. Background to Instrument Development**

The C-test was first developed in 1981 as a test of general language proficiency, and has been used mainly with children learning their first or second language, or with adults learning a foreign language (Grotjahn, Klein-Braley, & Raatz, 2002). Like the Cloze procedure it makes use of the reader’s ability to predict missing information from the surrounding context. However, in the C-Test the words in the text are only partly deleted. Research into the C-Test has provided evidence of its use and acceptability as a measure of general language

proficiency (Connelly, 1997; Eckes & Grotjahn, 2006; Grotjahn et al., 2002; Katona & Dornyei, 1993)

In its original form, the C-test deletion procedure followed the “rule of two” (i.e. beginning with the second word of the second sentence, the second half of every second word is deleted). Exceptions were made for proper names and numbers, but otherwise the rule was strictly applied (Raatz & Klein-Braley, 2002). Subsequent studies have seen a more flexible approach applied to test construction. Jafarpur (1999), after experimenting with alternative versions of the C-test using different deletion starting points and ratios, concluded that the rule of two offered no advantages over the other forms. Ishihara, Hiser & Okada (2003) used passages that were longer and more complex than those traditionally used in C-testing and reached similar conclusions. Deleting the last half of every other word from these long passages would have given far too many blanks to be filled in by the respondents, so the researchers therefore reduced the ratio of deletion to every fifth or sixth word. To test the validity of their revised format, they created three versions of the test with different deletion starting points. The scores on the modified C-Tests exhibited high reliability as well as high inter-version correlations. C-tests have most commonly been used for second language testing with the assumption that native speakers should be able to obtain near perfect scores: 90% or higher. Klein-Braley & Raatz (1984, p.144) state that a test administered at the end of a language teaching unit should have an average score of “80%, 90% even 100%”.

The Sentence Verification Technique (SVT) was introduced in 1979 as a method for measuring reading comprehension. SVT is based on a psycholinguistic theory of language comprehension whereby a successful reader constructs meaning by preserving a memory representation of what they have read (Royer, 2001). Thus, cognitive understanding is premised on short-term memory. Its initial trial, with primary-age school children, showed encouraging results (Royer, Hastings, & Hook, 1979). Subsequent research has shown that SVT tests have good reliability and validity as tests of reading comprehension in a variety of situations (Royer, Carlo, Dufresne, & Mestre, 1996; Royer, Lynch, Hambleton, & Bulgareli, 1984; Royer, Marchant, Sinatra, & Lovejoy, 1990; Royer, Tirre, Sinatra, & Greene, 1989). SVT tests can be used with virtually any text-based reading material.

The SVT adopts a structured approach to the construction of questions for a given passage or text. Each question (or item) is framed as one of four types of sentence. We take the definitions of the altered sentences from Royer (2004 p.5) and Royer (2004, p.67)<sup>1</sup>:

- an original sentence (a sentence presented exactly as it appears in the passage);
- A paraphrase sentence which is constructed by changing as many words as possible in an original sentence without altering the meaning of the sentence. The general guidelines for the construction of paraphrase items are: 1) change as many words as possible; 2) make sure the paraphrase has the same meaning as the original sentence; and 3) make sure the meaning of the paraphrase sentence fits with the meaning of the passage as a whole.
- A meaning change sentence which is constructed by changing one or two words in the sentence so that the meaning of the sentence is altered. The general guidelines for constructing meaning change items are: 1) substitute one or two words in the original sentence so that the meaning of the new sentence is distinctly different from the original; 2) make sure the meaning of the sentence is inconsistent with the meaning of the passage as a whole; and 3) make sure the meaning of the test sentence is not bizarre in any way.
- A distractor sentence which is a sentence that has a syntactic structure that is similar to a sentence in the passage and is consistent with the overall theme of the text passage, but is unrelated in meaning to any sentence that appeared in the passage. The general guidelines for writing distractors are: 1) make sure the distractor is roughly comparable to original sentences in terms of length, syntactic structure, and vocabulary level; 2) make sure the distractor sentence is consistent with the thematic content of the passage; and 3) make sure the distractor sentence is different in meaning from any original sentence in the passage.

Royer (2004, p.5) observes that meaning change sentences are the most difficult to write, since the nature of the change may vary between 'obvious' and 'subtle'. Given the relative sophistication of the target audience in this paper, the authors preferred 'subtle' variety of 'meaning change' questions. The test is generally balanced in that the reader is presented with an equal number of items of the four sentence types. For each item presented, the reader must decide whether it has the same meaning as, or a different meaning from, a sentence in the text. This is usually presented as a Yes/No alternative. The SVT was developed for use

with short passages (12 sentences) and traditionally has 12 or 16 questions per passage (Royer, 2001). However, it has also been used with much longer passages by first sampling important sentences from the text and then constructing SVT items from these sentences (Royer et al., 1989).

Some examples of applications of the SVT test are to assess university students' comprehension of textbook materials (Durwin & Sherman, 2008) and to check detailed understanding of text from GP-authored patient information leaflets (Newton et al. 1998). In this latter study of UK Patient Information Leaflets, the YES/NO statements were printed on the back of the test passage and respondents were required to remember what they had just read. Newton et al (1998) also presented the statements in the same sequence as the sentences in the original passage. This presentational format deviated from, Royer's (the deviser of the test) earlier advocacy of a random arrangement of the items (with the restriction that test sentences assessing the first half of the original passage appear first). In practice, this becomes far too confusing. We, consequently, adopted, Newton et al's (1998) research design so that the test items follow the order of the original statements.

Overall when interpreting SVT tests, Royer (2001) states that *if tests are based on materials appropriate for the population being tested*, then the average examinee gets about 75% correct. Newton et al (1998) tested two versions of a text – an original and a simplified version with shorter words and sentences and fewer technical terms. The average score for the original was 66% (range: 54% to 88%) while the average score for the simplified text was 74% (range: 46% to 92%). This substantial improvement in scores was put down as evidence of a substantial improvement in reading ease. Poor comprehenders score 70% or below while good comprehenders score 80% and above.

A variation on the SVT, the Meaning Identification Technique (MIT), was developed by Marchant, Royer and Greene (1988) in an attempt to increase the reliability and validity of the SVT test. They had found that the discriminatory power of 'original' and 'distractor' sentences was much lower than that of 'meaning change' and 'paraphrase' sentences, hence they proposed an alternative form of the SVT, using 'paraphrases' and 'meaning change paraphrases'. 'Meaning change paraphrases' are formed by first writing a paraphrase of the sentence and then changing one or two words in the sentence so as to alter the meaning. Both types of test sentences use different words from those used in the passage, but only one - the

paraphrase sentence - has the same meaning. When interpreting these tests the working assumption for the SVT is that the average examinee gets about 75%, with poor comprehenders scoring up to 70% and good comprehenders scoring more than 80%. Marchant et al. (1988) suggest that MIT scores may be expected to be the same, or slightly lower than the SVT equivalent.

#### **4. Research Method**

We devised four different tests to measure comprehension using a common documentary base. Two of the tests we compiled were Cloze related: a conventional CLOZE test, and a simpler variant, the C-test, and two were ‘question and answer’ comprehension tests (the MIT and SVT procedures). The four tests were based on a real-world, narrative from an Australian listed company’s annual report. An experiment was designed to match readers’ performance with one of each kind of test (i.e., one CLOZE-based (Cloze or C-test), with one comprehension test (SVT or MIT). The tests were administered to university students. Highly correlated performance outcomes should indicate that the two instruments are measuring a common factor.

##### **4.1 Development of the Instrument**

###### **Test Construction**

The four tests were constructed using a sample Chief Executive Officer’s (CEO’s) statement taken from the 2006 Annual Report of the ASX-listed company, Newsat Limited (NWT). NewSat Limited is an independently operated satellite communications company, with headquarters in Melbourne, Australia. It provides communication services, including high-speed internet, video conferencing and VOIP telephony to Australia/Oceania, Asia, the Middle East and Africa. NewSat was chosen as a representative company for this study largely based on the length of its CEO statement. This was because as it was comparatively short, the statement could be used in its entirety, without the necessity of choosing random selections of text from the whole content.

CEO statements, although conventionally supplied by companies, are discretionary in nature. They normally provide an overview of a company's operations and activities. The 294 word accounting narrative by Newsat Limited dealt with seven main issues: the purchase of the intellectual capital of Mediapart; the sale of Multie Technology Distribution Pty Ltd.; the sale of Airworks Media Pty Ltd.; the change of the company's name to Newsat; the financial results of the company; its international growth; and a thank you to employees. We follow both Royer (2004, p.11) and Royer et al. (1989) in selecting relatively short passages (i.e. 294 words) for SVT/MIT tests, while being long enough to permit reliable assessment. The passage is written in the past tense by the Chief Executive, and has a Flesch reading ease score of 37.8, suggesting that the narrative is 'difficult', but within the 30-50 range normally associated with CEO statements.

The four test types are discussed below. They are shown in the Appendices to the paper. In each case the subjects were given a set of written instructions. The instructions were also explained verbally by the test instructor. The tests were supervised under exam conditions and the tests took approximately twenty minutes each. The respondents filled in an ethical compliance form signalling that their participation was voluntary and that they could, if they wished, withdraw from it at any time.<sup>2</sup>

#### Test 1: CLOZE Test (see Appendix 1)

This test involves replacing every tenth word of the statement with a dotted line and then requiring the respondent to fill in the missing words. The dotted lines were of standard length (ten dots). For the chosen statement, this resulted in 28 items for the respondent to complete. If we take the first two sentences as illustration, "Your Company has achieved a great deal this year. .... concluded a hugely successful capital raising to purchase the ..... property and Mediaport assets of NSS". Then, the first blank represents 'it' and the second represents 'intellectual'.

#### Test 2: C-Test (see Appendix 2)

This test was constructed in a similar way to that of Test 1 and, indeed, is a variant of the standard Cloze procedure. However, in this case the second half of every tenth word in the statement was replaced by a dashed line (with a standard six dashes). Where the word has an odd number of letters then the larger part of the word is deleted. The respondent

was then required to restore the missing parts of the words. The same words were deleted as in the Cloze Procedure. Prima facie, it would be expected that this test would be simpler than Test 1. Again if we take the first two sentences as illustration, “Your Company has achieved a great deal this year. I----- concluded a hugely successful capital raising to purchase the intell----- property and Mediaport assets of NSS”. Again the first truncated word is ‘it’ and the second is ‘intellectual’. For both Tests 1 and 2 numerical and date expressions were ignored, and were not treated as ‘words’ for the deletion process.

### Test 3: Meaning Identification Technique (MIT) test (see Appendix 3)

This takes the form of a comprehension test, where the respondent must read the CEO’s statement and then answer Yes/No to a series of items, without referring back to the text. Each item takes the form of either a paraphrase of a sentence in the original text, or a paraphrased sentence from the original text where one or two words have been changed in order to change the meaning of the sentence. No sentences were thus identical. The test contains twelve items: six paraphrase items and six meaning change paraphrase items. To illustrate, the original second sentence reads: “It concluded a hugely successful capital raising to purchase the intellectual property and Mediaport assets of NSS”. This has been paraphrased as “Multiimedia acquired the media port assets of NSS”. The meaning has remained the same and this is a paraphrase item. Then, the original third sentence is “Those assets are debt free and reside on 2 acres in Adelaide and Perth”. This has been paraphrased to “Those assets include unsettled debt and reside on 2 acres in Adelaide and Perth”. The meaning has thus changed from the assets being debt free to having unsettled debt. This is thus a meaning change paraphrase item. The test items follow the order of the original sentences in the text. For instance, an item based on the first sentence in the statement text appears before an item based on the fifth sentence. The respondent must answer Yes/No according to whether they think the meaning of the item is the same as a sentence in the text. So the correct response will be “Yes” for paraphrase items and “No” for meaning change items. There were six yes answers and six no answers. Thus, random guessing would, on average, score 50%.

### Test 4: Sentence Verification Technique (SVT) test (see Appendix 4)

This test is a variant of test 3, the difference being that in this case there are four types of test item. In addition to paraphrase items and meaning change items, there are also

original (unchanged) sentences and “distractor” sentences. Thus the original sixth sentence is repeated verbatim “The Company sold Multie Technology Distribution Pty Ltd, its hardware distribution operation”. A distractor sentence is one which is constructed to be consistent in theme to an original statement in the passage, but is unrelated in meaning to any sentence that appears in the passage. Thus, the sentence “This acquisition will make the company one of the largest media port operators and satellite providers in the southern hemisphere” is included. This sentence looks superficially similar but has no direct relation to any sentence which actually appeared in the text. Four items of each type are included, giving a total of sixteen questions as against twelve in the MIT test. The correct response will be “Yes” for original and paraphrase items, and “No” for meaning change and distractor items. Given that there were eight Yes answers and eight No answers, as with the SVT test random guessing would, on average, score 50%. Royer (2004, p.9) suggests that 12 sentence or 16 sentence SVT tests are the norm. Our use of a 20 sentence passage for testing purposes in this study would therefore appear to be reasonable.

### **Test Administration and Scoring**

The above tests were given to a class of university students. These students’ first language was English.<sup>3</sup> These students were studying a final year accounting course on a UK Business Degree.<sup>4</sup> They had a good knowledge of accounting. They were thus considered reasonable subjects for a test of materials contained in an annual report. In research terms, although not highly sophisticated users, they can be classed as reasonably informed users given their knowledge of financial reporting. A number of authors, most notably Ashton and Kramer (1980), Abdolmohammadi and Wright (1987), Brownell (1995) and Liyanarachchi (2007) have advocated the use of student surrogates in experimental settings such as this, where their performance is unlikely to differ radically from those of ‘real-world’ decision makers.

Each student was given either a CLOZE or a C-test, followed by one of the comprehension tests (either MIT or SVT). The order of administration was important because of the short-term memory effect of reading the full statement. In a comprehension test the student must first read the full (unadulterated) passage, and so a subsequent Cloze or C-test would become an easier exercise to complete. This facilitates four possible experimental groups and given



the number of available subjects sitting the tests was forty four, it was decided to limit the experiment to two groups, A and B.<sup>5</sup>

The four tests are given in Appendices 1, 2, 3 and 4. For each test a detailed set of instructions was given. The students were told that “the following task is a comprehension test based on the Chairman’s Statement from a company’s Annual Report. The test measures how easy the statement is to understand”. For the Cloze test they were told “every tenth word has been replaced by a dotted line” and that “the length of each dotted line is the same and is not related to the length of the missing word. The task is to fill in the words which you think have been deleted”. They were then given an example with five deletions followed by the answer. The test passage and the answer passage are given in Appendix 1A and 1B. For the C-test a similar set of instructions was given. In this case, the students were told that “the second half of every tenth word has been replaced by a dashed line. If a word has an odd number of letters then the larger part of the word has been deleted, e.g. proud becomes pr----- . Please note that the length of each dashed line is the same and is not related to the length of the missing part of the word. The task is to restore the missing parts by writing in the letters which you think have been deleted.” An example with six deletions was then given, followed by the answer. The test question and the test answer are then given in Appendices 2A and 2B. The MIT and SVT instructions and reading passage are given in Appendix 3A and Appendix 3B. These are followed in Appendix 3C and 3D by the MIT questions and answers and in 4A and 4B by the SVT questions and answers.

Group A (23 respondents) were given the CLOZE test and the MIT test while Group B (21 respondents) were given the C-test and the SVT test. In this way each of the four tests (CLOZE, MIT, C and SVT) was used once. The time allowed for the two tests was 30 to 40 minutes and the test was administered under exam conditions. Tests 2-4 were scored by allocating one mark to each correct response. Thus, test 2 (C-test) had a total possible score (TPS) of 28, Test 3 (MIT) a TPS of 12 and test 4 (SVT) a TPS of 16. Test 1 was handled slightly differently. This is because in this test, there is the problem of whether or not to count synonyms. The two main approaches in the literature are to either award a synonym a whole mark, or no marks at all. We preferred to adopt a compromise position by allocating one mark where the student correctly identified the missing word and half a mark where the student’s response was a synonym of the missing word. The maximum score was 28. The research literature (e.g., Miller & Coleman, 1967) suggests that it does not matter whether or

not synonyms are counted since the two outcomes are so highly correlated. However, our experience suggests that either of these extremes can upset the respondents (i.e., if a respondent gets the *right* answer, but somebody else get full marks for a synonym; or if a respondent puts down what s/he perceives as a perfectly acceptable alternative, but get no credit for it). We therefore chose this marking procedure to alleviate the wrath of the participants!

## 5. Results and Interpretation

In Table 1, we present a summary of the percentage test scores. It must be remembered that test groups 1 (Cloze) and 3 (MIT) were paired as were tests 2 (C-test) and 4 (SVT). In each case, we present the descriptive data: the number of subjects (n), the mean% scores, the standard deviation, the median% scores, the minimum and maximum scores achieved and the total possible raw score (TPS) –i.e., number of correct responses achievable.

Test	n	Mean	Std. Dev.	Median	Min	Max	TPS
1: CLOZE	23	57.3	11.007	58.9	33.9	73.2	28
2: C-test	21	81.8	12.873	82.1	39.3	100.0	28
3: MIT	23	73.9	13.602	75.0	50.0	91.7	12
4: SVT	21	70.5	10.485	68.7	50.0	87.5	16

It seems sensible, first to compare the Cloze and C-tests. As expected, given the structure of the tests, the scores in the Cloze test were significantly lower (mean 57.3) than those in the C-test (81.8) [unpaired t-test:  $t=-6.8045$   $p<0.01$ ]. This reflects the fact that the Cloze tests measure whole word deletions, while the c-test measures only partial deletions.

Turning to a comparison of the MIT and SVT results, then it should be borne in mind that a random score would have been 50%. In fact, this was the minimum score for both tests. Some subjects, therefore, performed no better than on an average random basis. Fortunately, no student scored worse than random! Scores for the MIT test were slightly higher (73.9) than those for the SVT tests (70.5), but the difference was not statistically significant ( $t=0.9159$ ,  $p=0.3649$ ). This slight increase in mean scores may simply reflect the fact that the MIT test is simpler, with fewer alternatives, than the SVT test.<sup>6</sup>

The interpretation of these scores, as we discuss in the next section, is fraught with difficulty. However, we will use the prior guidance as a referential benchmark score. For the Cloze test, we use 57% as the independent level at which readers will find a passage understandable. In our study, the mean was 57% and 8 out of 23 students (35%) scored below 57%. This was above the 10% threshold that Newton et al (1998) suggested was necessary for a passage to be easy to read. For the C-tests too the scores which were, on average 81.8, and thus below 90% indicating that the text was difficult.

The SVT tests average score was 70.5%, and again this was below the expected average of 75% correct. There is very little literature available to assist with the interpretation of MIT scores – although Marchant, Royer and Greene (1988) argue that MIT scores may be lower than the average recommended for the SVT scores. Thus a mean score of 70.5% could indicate text that is above average in difficulty.

Overall, taken together, the tests seem to indicate a slightly below average comprehension level, and the results across tests look quite consistent. The only slightly surprising feature is the higher scores on the MIT compared with the SVT, although the difference was shown not to be statistically significant.

For Group A, the outcomes for Tests 1 and 3 (Cloze and MIT) were not significantly correlated ( $r=0.290$ ,  $p=0.179$ ). For Group B correlation between Test 2 and Test 4 (C-test and SVT) was better ( $r=0.385$ ,  $p=0.085$ ). However, neither relationship was significant at the 5% level. This suggests that the Cloze-based (Cloze and C-test) and question and answer (MIT and SVT) tests measure different things.

## **Discussion**

The four tests needed to be constructed, administered and interpreted. We, therefore, evaluate them against these criteria below, using our own experience, and also drawing, where appropriate, upon the prior literature. We also compare them to more traditional reading comprehension tests.

### **i) Construction**

We found both the Cloze Procedure and C-Test were very easy to construct. The MIT and SVT tests by their very nature we found more difficult to construct. A reading comprehension test gives a quantitative measure of a text's comprehensibility and is therefore more directly linked to what de Jong & Schellens (1997) term "functional communication". However, a lot of time and expertise is required to formulate good comprehension questions about a text (de Jong & Schellens, 1997). Weir comments that comprehension tests are expensive and difficult to prepare; writers need to be specially trained; and construction is time-consuming. When compared to the traditional comprehension test, however, we found the MIT and SVT tests, had some constructional advantages. Indeed, investigators have been able to construct SVT tests using Royer's descriptions of the development process, without the need for any special training (Royer, 2004). We found that the two tests were relatively easily and quickly constructed given their formulaic and standardised nature and relatively objective presentational format. If anything, we found, MIT seems superior in terms of ease of construction. Although, we acknowledge that in all tests of this kind we suspect that an extremely detailed examination of the questions employed would allow for possible improvements to be made.

## ii) Administration

The Cloze and C-tests were economical to administer. One advantage that the C-test had over Cloze was that it could be objectively scored as there was only one possible answer for an item (Katona & Dornyei, 1993; Weir, 1990). We found scoring the cloze test more difficult because of the synonyms. Katona & Dornyei (1993) claim that it is a less frustrating task for learners than the Cloze test, and that students actually enjoy taking the tests. The C-test is usually easier, especially for native language speakers, than a Cloze test, and C-tests may be less frustrating for respondents based on this factor. However, Jafarpur (1995, p. 87) disagrees; he claims that C-tests, with their heavily mutilated appearance, are "*irritating and unacceptable*" to test subjects, and concludes that its face validity is low since subjects think of it more as a type of word puzzle than a test of language ability. We found the MIT and SVT tests relatively easy to mark and score. They all had one correct answer and, unlike Cloze, do not need to take synonyms into account.

### iii) Interpretation

The interpretation of the four tests presented us with far more problems. This is in terms of what the actual test is measuring: is it, for example, reading comprehension or something else, and what do the actual scores mean? The Cloze Procedure has been argued to measure redundancy rather than reading comprehension per se and this is true of the C-Test. This is because the Cloze procedure involves puzzle solving and cognitive processes that may or may not be synonymous with understandability.

There has been considerable discussion about the interpretation of the level of scores. For the Cloze procedure, the instructional level is generally considered to be 57%, but Jones (1998) shows the flexibility of this quantitative benchmark. For the SVT tests and MIT tests, there is also considerable variability. For the SVT test, the working assumption is that the average respondent examined gets about 75% with poor comprehenders scoring below 70% and good comprehenders scoring more than 80%. Royer (2004, p.19) suggests an SVT performance range of: >80% (good); 75% (average); <70% poor, but recognises (p.24) that SVT test scores will be sensitive to text difficulty, and particularly to understandability of questions. This is in line with scores from normal comprehension tests. It also prima facie makes sense from our results where we might have expected that students would find the accounting texts quite difficult to read as they are very technical. There is very little in the literature about MIT tests, particularly when it comes to the interpretation of test scores. The implication in the paper by Marchant, Royer and Greene (1988) is that the average score should decrease when original and paraphrase items are omitted because examinees tend to score higher on these types of items. Interestingly, this is not the case with our results where the average MIT score is actually higher than the SVT score. This may be because the limited choice of the MIT test allows students to focus more, or that the SVT distractor items are just too distracting! Royer (2004, p.7) suggests that the bulk of the discriminatory power of the SVT test lies in the 'paraphrase' and 'meaning change' items. Since these are the only items in the MVT test Marchant et al. (1988) suggest that the latter has more reliability and validity. As the results from the SVT and MIT tests were, however, very similar and statistically it was not possible to distinguish between them, they would prima facie appear to be measuring the same construct. As MIT is easier to construct than this measure seems preferable. Given the similarity between normal comprehension tests, SVT and MIT, we decided that a benchmark of 75% would be appropriate when attempting to assess the degree of understanding.

## **Conclusions and Limitations**

The psycholinguistic literature has long questioned the efficacy of the Cloze procedure as a measure of understandability. However, until now this criticism has failed to permeate the accounting literature. This study has tested four potential understandability measures: the Cloze procedure, the C-test, MIT and SVT. Only the Cloze procedure has been studied previously in an accounting context. The four tests were administered under a controlled experimental context to 44 UK undergraduate students. Overall, the students' understanding of the passages, based on conventional interpretation of each of the measures, appeared to be just below average.

We evaluated these tests against their ease of construction, administration and interpretation. In general all four tests were relatively simple to design with the Cloze procedure and C-test being the easiest, requiring only mechanical manipulation of the text, rather than the generation of original test questions as for SVT/MIT. On administration, the SVT, MIT and C-test were once again the easiest to administer. The Cloze procedure was slightly harder given problems of scoring synonyms. Interpretation is more problematic, since the Cloze procedure's benchmark has been criticised while little is known about the C-test. For the MIT and SVT test, there appears to be a consensus that a 75% benchmark, as in multiple choice tests is reasonable. However, Royer (2004) recognises that SVT cut-offs are sensitive to the difficulty of the text, and particularly to the complexity of the questions, so that, as in this case, a difficult text combined with a sophisticated target audience makes rigid adherence to a 75% benchmark unwise.

Earlier studies (e.g., Smith & Taffler, 1992b; Jones 1988) indicate that readability and understandability are different concepts. The statistical tests reported here further suggest that Cloze/C-test instruments measure something different to that of the MIT/SVT tests. If the latter are measuring 'understandability' then the former are not. Earlier studies (e.g., Smith & Taffler, 1992b; Jones 1988) suggest that readability and understandability are different concepts. Indeed, it is unclear here exactly what the Cloze test does measure! It is likely, however, to measure redundancy and is closely connected with the predictability of text. Unfortunately, this does not appear to correspond with understandability, as borne out by the

insignificant correlation measures between the Cloze type tests and the SVT/MIT comprehension measures. The SVT and MIT tests, by contrast, do more persuasively measure understandability.

These findings overall do show that these four tests are easy to construct and administer. Moreover, the SVT and MIT tests also have the potential to evaluate the comprehension of accounting narratives. However, the findings are limited to one group of respondents, comprising just 44 undergraduate degree accounting students, so they may not be generalisable. Also there is the potential for case specificity, since only one company narrative has been used. There is obviously need for further research that will test financial communication in a variety of contexts.

Given that it appears the Cloze procedure and C-test are not strictly measures of understandability, we would particularly encourage accounting researchers to continue experimentation with the MIT and SVT tests. If it is accepted that MIT and SVT do, at least to some extent, surrogate understandability, then the obvious question is which test is superior. Trials conducted by Marchant et al.(1988) indicated that the MIT procedure was more reliable and superior in validity to traditional SVT tests, though the MIT procedure was not superior in all circumstances. Marchant et al., (1988, p. 833) suggest that "*investigators who are interested in using SVT tests as measures of comprehension ... should consider using the MIT procedure rather than the SVT procedure*". This view is reinforced by Fisher, Fox, & Wood (1999, p. 56). In their study of reading difficulties with Scottish school children, they used the MIT technique as a listening comprehension assessment tool. However, they discounted the SVT as "cumbersome and difficult to develop", the MIT was thought to be more practical, as well as demonstrating greater validity and reliability.

The MIT test seems to be the most supported in the literature: it is relatively easy to construct, relatively easy to mark and score, and a working benchmark of 75% would seem to approximate a benchmark for understanding; we would recommend that further testing of the MIT test be conducted. A larger scale test could test the effectiveness of the MIT test, for example, against the Cloze Procedure and SVT test, but also in experimental settings. These might involve the testing of the understandability of annual reports or institutional accounting pronouncements such as those by the International Accounting Standards Board. Two examples might be, first, comparing the relative understandability of the accounting

narratives of annual reports of profitable and unprofitable companies, and second comparing the relative understandability of different language (i.e., English, French or German) versions of the IASB’s conceptual framework or standards. Prima facie, the MIT test would be particularly useful in such a comparative multi-language setting.

## Footnotes

<sup>1</sup> We are grateful to an anonymous reviewer for detailed comments on the construction of the test. This reviewer felt that our distractor sentences were not comparable to the original sentences because they did not a similar structure to the one in the passage. We believe, however, that our distractors did comply with Royer’s other two guidelines, they were consistent with the thematic content and were different in meaning from the original sentence( Royer 2006, p.6). We, therefore, believe they comply with the spirit of Royer’s distractors. The distractors were only relevant to the SVT question and not to any of the other three tests.

<sup>2</sup> The study was thus carried out in accordance with the University’s standard research ethics procedures.

<sup>3</sup> After the test, it was pointed out that there were two minor errors in the tests Mediaport and Multie were spelt inconsistently. Although unfortunate it is not considered that these minor inconsistencies would have caused serious problems.

<sup>4</sup> There is thus a potential mismatch between using an Australian report and a set of UK experimentees. However, given the universal nature of English this is not believed to have caused any particular confounding effects or internal validity problems.

<sup>5</sup> There is thus a possibility that the cognitive processing of the Cloze and C-test, may have been slightly different and affected their scores on the comprehension tests (SVT and MIT). However, we judged this effect to be minimal given the similarity in cognitive processes for the two tests and the fact both groups had the same full passage for the comprehension test.

<sup>6</sup> One of the reviewers queried whether the four meaning changes in the SVT test sufficiently differentiated the new sentence from the original. In order to test this we carried out a follow up validation study with 27 final year students with a comparable knowledge of accounting terms and concepts to our test cohort. The respondents were asked to clarify the difference between terms in the original statement and question change statement. The results are shown below.

SVT Question	Original Statement	Question Change	Recognition of Difference in Indicated Responses
Q2.	unsettled debt	debt-free	24/25 = 96%
Q7.	diversify	divest	19/23 = 96%
Q13.	gains	profits	21/25 = 84%
Q14.	Eastern Europe	Middle East	27/27 = 100%

The results of the investigation demonstrated an overwhelming recognition of differences among the terms, sufficient to justify their continued use as ‘meaning change’ questions in the SVT and MIT test instruments.



## References

- Abdolmohammadi, M. and Wright, A. (1987) "An Examination of the Effects of Experience and Task Complexity on Audit Judgments", *The Accounting Review*, January, 1-13.
- Abraham, R.G. and Chapelle, C.A (1992), "The meaning of cloze test scores: an item difficulty perspective", *The Modern Language Journal*, 76, iv, 468-79
- Adelberg, A.H. (1983), "The accounting syntactic complexity formula: a new instrument for predicting the readability of selected accounting communications", *Accounting and Business Research*, Summer, 162-75
- Adelberg, A.H. and Razeq, J.R. (1984) "The Cloze Procedure: A Methodology for Determining the Understandability of Accounting Textbooks", *Accounting Review*, 59 (1), 109-122.
- Aerts, W. (2010) "Self-serving tendencies in retrospective reasoning in accounting narratives: Does a capital market environment matter?" Working paper, University of Antwerp.
- Arthur Andersen (2001) *Spice up the Story: A Survey of Narrative Reporting in Annual Reports*, Arthur Andersen, London.
- Ashton, R.H. and Kramer, S.S. (1980) "Students as Surrogates in Behavioral Accounting Research: Some Evidence", *Journal of Accounting Research*, 18, 1-15.
- Bailey, M. and Harrison, C. (1984), "Cloze procedure as a measure of reading comprehension: the effects of variations in subject matter on textual redundancy", *Human Learning*, 3, 185-201.
- Bormuth, J.R. (1967), "Comparable cloze and multiple choice comprehension test scores", *Journal of Reading*, 10, 291-99
- Bormuth, J.R. (1968) "Cloze Test Readability: Criterion reference Scores", *Journal of Educational Measurement*, 5 (3), 189-96.
- Brennan, N.M., Guillamon-Saorin, E. and Pierce, A. (2009) "Impression management: developing and illustrating a scheme of analysis for narrative disclosures – a methodological note", *Accounting, Auditing and Accountability Journal*, 22 (5): 789-832.
- Brownell, P. (1995) *Research Methods in Management Accounting*, Melbourne, Coopers & Lybrand.
- Bryan, S.H. (1997) "Incremental information content of required disclosures contained in management discussion and analysis", *The Accounting Review*, 72 (2), 285-301.
- Chall, J.S. (1958), "Readability – an appraisal of research and application. Bureau of Educational Research. Ohio, Ohio State University Press.
- Clatworthy, M.A, and Jones, M.J. (2001), "The effect of thematic structure on the variability of annual report variability of annual report readability", *Accounting, Auditing and Accountability Journal*, 14, (31), 311–326.
- Cohen, J.H. (1975), "The effect of content area material on cloze test performance", *Journal of Reading*, December, 247-50.
- Connelly, M. (1997). Using C-tests in English with post-graduate students. *English for Specific Purposes*, 16 (2), 139-150.

- Courtis, J.K. (1986), "An investigation into annual report readability and corporate risk return relationships", *Accounting and Business Research*, Autumn, 285-194.
- Courtis, J.K. (1996) "Disclosure Redundancy in Annual Reports", *Accountability and Performance*, 2 (3) ; 1-16.
- Courtis, J.K. (1998), "Annual report readability variability: Tests of the obfuscation hypothesis", *Accounting, Auditing and Accountability Journal*, 11, (41), 459-475.
- Courtis, J.K. (2004) "Corporate report Obfuscation: Artefact or Phenomenon?" *British Accounting Review*, 36 (3), 291-312.
- de Jong, M., & Schellens, P. J. (1997). Reader-Focused Text Evaluation: An Overview of Goals and Methods. *Journal of Business and Technical Communication*, 11(4), 402-432.
- Durwin, C. C., & Sherman, W. M. (2008). Does Choice of College Textbook Make a Difference in Students' Comprehension? *College Teaching*, 56 (1), 28-34.
- Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing*, 23 (3), 290-325.
- Entin, E.B. and Klare, G.R. (1978), "Some inter-relationships of readability, cloze and multiple choice scores on a reading comprehension test", *Journal of Reading Behaviour*, 10, (4), 417-36
- Fisher, D., Fox, D., & Wood, K. (1999). The identification of specific reading difficulties through assessment of listening comprehension skills. *Educational and Child Psychology*, 16 (1), 54-59. Retrieved from <http://www.bps.org.uk/decp/educational-and-child-psychology/educandchildpsych.cfm>
- Gough, P.B. (1972), "One second of reading", in Singer, H. and Ruddell, R.B. (Eds), *Language by Ear and Eye*, International Reading Association, Newark, DE, 47-94
- Grotjahn, R., Klein-Braley, C., & Raatz, U. (2002). C-Tests: An Overview. In J. A. Coleman, R. Grotjahn & U. Raatz (Eds.), *University Language Testing and the C-Test* (pp. 93-114). Bochum, Germany: AKS-Verlag.
- Ishihara, K., Hiser, E., & Okada, T. (2003). Modifying C-Test for Practical Purposes. *Doshisha Studies in Language and Culture*, 5(4), 539-568. Retrieved from <http://elib.doshisha.ac.jp/english/kiyo.html>
- Jafarpur, A. (1995). Is C-testing superior to cloze? *Language Testing*, 12 (2), 194-216.
- Jafarpur, A. (1999). What's Magical About the Rule-of Two for Constructing C-Tests? *RELC Journal*, 30 (2), 86-100.
- Jones, M.J. (1988) "A Longitudinal Study of the Readability of the Chairman's Narratives in the Corporate Reports of a UK Company", *Accounting and Business Research*, 18 (72), 297-305.
- Jones, M.J. and Shoemaker, P.A (1994), "Accounting narratives: a review of empirical studies of content and readability", *Journal of Accounting Literature*, 13, 142-84.
- Jones, M.J. (1997), "Critical appraisal of the Cloze procedure's use in the accounting domain", *Accounting, Auditing and Accountability Journal*, 10 (1), 105-128.
- Katona, L., & Dornyei, Z. (1993). The C-Test: A Teacher Friendly Way to test Language Proficiency. *English Teaching Forum*, 31 (1), 35. Retrieved from <http://eca.state.gov/forum/vols/vol31/no1/p35.htm>

- Kintsch, W. and Miller, J.R. (1984), "Readability: a view from cognitive psychology", in Flood, J. (Ed.), *Understanding Reading Comprehension: Cognition, Language, and the Structure of Prose*, International Reading Association, Newark, DE, 220-32
- Kintsch, W. and Vipond, D (1979) "Reading comprehension and readability in educational practice and psychological theory" in L.G. Nilsson (ed.) *Perspectives on Memory Research* (pp.329-366) Hillsdale NJ: Erlbaum.
- Klare, G.R. (1963). *'The Measurement of Readability'*, Iowa State University press, Ames, Iowa
- Klein-Braley, C., and Raatz, U. (1984). A survey of research on the C-Test1. *Language Testing*, 1 (2), 134-146.
- Liyanarachchi, G.A. (2007) "Feasibility of using student subjects in accounting experiments: a review", *Pacific Accounting Review*, 19 (1), 47-67.
- McNamara, D.S., Kintsch, E., Butler-Songer, N. and Kintsch, W. (1996) "Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text", *Cognition and Instruction*, 14, 1-43.
- Marchant, H. G., III, Royer, J. M., & Greene, B. A. (1988). Superior Reliability and Validity for a New Form of the Sentence Verification Technique for Measuring Comprehension. *Educational and Psychological Measurement*, 48 (3), 827-834.
- Merkel-Davies, D.M. and Brennan, N. (2007), "Discretionary disclosure strategies in corporate narratives: Incremental Information or Impression Management, *Journal of Accounting Literature*, 26, 116-195.
- Miller, G.R. and Coleman, E.B. (1967) "A Set of Thirty-Six Prose Passages Calibrated for Complexity", *Journal of Verbal Learning and Verbal Behaviour*, 6, 851-4.
- Newton, L., Newton, D., Clark, J., Kenny, T., Moseley, D., Purves, I., et al. (1998). Patient information leaflets: producing understandable PILs. *Journal of Information Science*, 24 (3), 167-181.
- Parker, L.D. (1982), "Corporate annual reporting: a mass communication perspective", *Accounting and Business Research*, Autumn, 279-286.
- Raabe, A., Stevens, K.C. and Stevens, W.P. (1984), "Tax textbook readability: an application of the cloze method", *Journal of American Taxation Association*, 6 (1), 66-73.
- Raatz, U., & Klein-Braley, C. (2002). Introduction to language Testing and to C-tests. In J. A. Coleman, R. Grotjahn & U. Raatz (Eds.), *University Language Testing and the C-Test* (pp. 75-91). Bochum, Germany: AKS-Verlag.
- Rankin, E. (1971), "Grade level interpretation of cloze readability scores", in Greene, F.B. (Ed.), *Reading: The Right to Participate*, National Reading Conference, Milwaukee, WI, 30-7.
- Rankin, E.F. and Culhane, J.Y. (1969), "Comparable cloze and multiple-choice comprehension test scores", *Journal of Reading*, December, 93-198
- Royer, J. M. (2001). Developing reading and listening comprehension tests based on the Sentence Verification Technique (SVT). *Journal of Adolescent & Adult Literacy*, 45 (1), 30-41.
- Royer, J. M. (2004). Uses for the Sentence Verification Technique for Measuring Language Comprehension. Unpublished manuscript. University of Massachusetts.

- Royer, J. M., Carlo, M. S., Dufresne, R., & Mestre, J. (1996). The Assessment of Levels of Domain Expertise while Reading. *Cognition and Instruction*, 14 (3), 373-408.
- Royer, J. M., Hastings, C. N., & Hook, C. (1979). A sentence verification technique for measuring reading comprehension. *Journal of Literacy Research*, 11 (4), 355 - 363.
- Royer, J. M., Lynch, D. J., Hambleton, R. K., & Bulgareli, C. (1984). Using the Sentence Verification Technique to Assess the Comprehension of Technical Text as a Function of Subject Matter Expertise. *American Educational Research Journal*, 21 (4), 839-869.
- Royer, J. M., Marchant, H. G., III, Sinatra, G. M., & Lovejoy, D. A. (1990). The Prediction of College Course Performance from Reading Comprehension Performance: Evidence for General and Specific Prediction Factors. *American Educational Research Journal*, 27 (1), 158-179.
- Royer, J. M., Tirre, W. C., Sinatra, G. M., & Greene, B. A. (1989). Assessment of On-Line Comprehension of Computer-Presented Text. *Journal of Educational Research*, 82 (6), 348-355.
- Shaffer, R.J., Stevens, K.T, and Stevens, W.P. (1993), "Assessing the readability of government standards: the cloze procedure", *Journal of Technical Writing and Communication*, 23, 259-67
- Shanahan, R., Kamil, M.L and Tobin, A.W. (1982), "Cloze as a measure of intersentential comprehension", *Reading Research Quarterly*, 2, 229-55.
- Smith, J.E. and Smith, N.P. (1972). 'Readability: A measure of the performance of the communication function of financial reporting', 46 ( 3), 552-61
- Smith, M. and Taffler, R.J. (1992a), "The chairman's statement and corporate financial performance", *Accounting and Finance*, 32 (2), 75-90.
- Smith, M. and Taffler, R.J. (1992b), "Readability and Understandability: Different Measures of the Textual Complexity of Accounting Narrative", *Accounting, Auditing and Accountability Journal*, 5 (4), 84-98.
- Smith, M. and Taffler, R.J. (2000), "The chairman's statement: A content analysis of discretionary narrative disclosures", *Accounting, Auditing and Accountability Journal*, 15 (4), 624-646.
- Sydserrff, R. and Weetman, P. (1999), "A texture index for evaluating accounting narratives: An alternative to readability formulae", *Accounting, Auditing and Accountability Journal*, 12, (4), 459-488
- Sydserrff, R. and Weetman, P. (2002), "Developments in content analysis: A transitivity index and diction scores", *Accounting, Auditing and Accountability Journal*, 15, (4), 523-545
- Taylor, W.L. (1953), "Cloze procedure: a new tool for measuring readability", *Journalism Quarterly*, 415-33.
- Weir, C. J. (1990). *Test Methods. In Communicative Language Testing* (pp. 42-85). New York: Prentice-Hall International.

## Appendix 1 (a): CLOZE test

### Chairman's Statement

Your Company has achieved a great deal this year.

..... concluded a hugely successful capital raising to purchase the ..... property and Mediaport assets of NSS. Those assets are ..... free and reside on 2 acres in Adelaide and Perth. ..... Mediaports support a world wide network of prestigious customers ..... boast long term profitable revenues. They are now the ..... of Multiimedia's future and will grow substantially as we ..... more customers, support more satellites and develop increasingly popular ..... applications.

The Company sold Multie Technology Distribution Pty Ltd, ..... hardware distribution operation. This was in line with our ..... to divest of non-core operations.

The Board has received ..... 29 September 2006 a firm offer for the sale of Airworks Media ..... Ltd. This offer is under consideration and the Board ..... to make an announcement shortly. A sale of Airworks ..... Pty Ltd will be a wonderful achievement for a ..... founded in 2003 with just one employee.

Shareholders have voted ..... change Multiimedia's name to NewSat and in October this ..... take effect. The new name is well known in ..... telecommunications space and better reflects the Company's core business.

..... year revenues for continuing divisions were up 142% and our ..... 18.6% down. Directors were satisfied with this trend and look ..... to further gains in 2006/2007.

International business spearheaded from the ..... East also grew. The Company supports customers from SE ..... to North Africa and has an aggressive growth strategy ..... Central and Northern Asia in 2006/2007.

Multiimedia's staff continue to ..... hard and professionally for all shareholders. Our structure is ..... and strategy firm. The Board are to be also ..... for their diligence and attention to detail. I look ..... to the New Year with enormous enthusiasm.

## Appendix 1(b): Cloze Test Solution

### Solution

Your Company has achieved a great deal this year.

IT concluded a hugely successful capital raising to purchase the INTELLECTUAL property and Mediaport assets of NSS. Those assets are DEBT free and reside on 2 acres in Adelaide and Perth. THE Mediaports support a world wide network of prestigious customers AND boast long term profitable revenues. They are now the BACKBONE of Multiimedia's future and will grow substantially as we ACQUIRE more customers, support more satellites and develop increasingly popular COMMUNICATION applications.

The Company sold Multie Technology Distribution Pty Ltd, ITS hardware distribution operation. This was in line with our STRATEGY to divest of non-core operations.

The Board has received ON 29 September 2006 a firm offer for the sale of Airworks Media PTY Ltd. This offer is under consideration and the Board EXPECTS to make an announcement shortly. A sale of Airworks MEDIA Pty Ltd will be a wonderful achievement for a BUSINESS founded in 2003 with just one employee.

Shareholders have voted TO change Multiimedia's name to NewSat and in October this WILL take effect. The new name is well known in AUSTRALIAN telecommunications space and better reflects the Company's core business.

THIS year revenues for continuing divisions were up 142% and our LOSS 18.6% down. Directors were satisfied with this trend and look FORWARD to further gains in 2006/2007.

International business spearheaded from the MIDDLE East also grew. The Company supports customers from SE ASIA to North Africa and has an aggressive growth strategy ENCOMPASSING Central and Northern Asia in 2006/2007.

Multiimedia's staff continue to WORK hard and professionally for all shareholders. Our structure is STABLE and strategy firm. The Board are to be also THANKED for their diligence and attention to detail. I look FORWARD to the New Year with enormous enthusiasm.

## Appendix 2 (a): C-test

### Chairman's Statement

Your Company has achieved a great deal this year.

I----- concluded a hugely successful capital raising to purchase the intell----- property and Mediaport assets of NSS. Those assets are de----- free and reside on 2 acres in Adelaide and Perth. T----- Mediaports support a world wide network of prestigious customers a----- boast long term profitable revenues. They are now the back----- of Multiimedia's future and will grow substantially as we acq----- more customers, support more satellites and develop increasingly popular commun----- applications.

The Company sold Multie Technology Distribution Pty Ltd, i----- hardware distribution operation This was in line with our stra----- to divest of non-core operations.

The Board has received o----- 29 September 2006 a firm offer for the sale of Airworks Media P----- Ltd. This offer is under consideration and the Board exp----- to make an announcement shortly. A sale of Airworks Me----- Pty Ltd will be a wonderful achievement for a busi----- founded in 2003 with just one employee.

Shareholders have voted t----- change Multiimedia's name to NewSat and in October this wi----- take effect. The new name is well known in Austr----- telecommunications space and better reflects the Company's core business.

Th----- year revenues for continuing divisions were up 142% and our lo----- 18.6% down. Directors were satisfied with this trend and look for----- to further gains in 2006/2007.

International business spearheaded from the Mid----- East also grew. The Company supports customers from SE As----- to North Africa and has an aggressive growth strategy encomp----- Central and Northern Asia in 2006/2007.

Multiimedia's staff continue to wo----- hard and professionally for all shareholders. Our structure is sta----- and strategy firm. The Board are to be also tha----- for their diligence and attention to detail. I look for----- to the New Year with enormous enthusiasm.

## Appendix 2(b): C-test Solution

### Solution

Your Company has achieved a great deal this year.

IT concluded a hugely successful capital raising to purchase the intellectual property and Mediaport assets of NSS. Those assets are deBT free and reside on 2 acres in Adelaide and Perth. THE Mediaports support a world wide network of prestigious customers AND boast long term profitable revenues. They are now the backBONE of Multiimedia's future and will grow substantially as we acQUIRE more customers, support more satellites and develop increasingly popular communication applications.

The Company sold Multie Technology Distribution Pty Ltd, iTS hardware distribution operation This was in line with our strATEGY to divest of non-core operations.

The Board has received oN 29 September 2006 a firm offer for the sale of Airworks Media PTY Ltd. This offer is under consideration and the Board expects to make an announcement shortly. A sale of Airworks MeDIA Pty Ltd will be a wonderful achievement for a busiNESS founded in 2003 with just one employee

Shareholders have voted to change Multiimedia's name to NewSat and in October this will take effect. The new name is well known in AustrALIAN telecommunications space and better reflects the Company's core business.

ThIS year revenues for continuing divisions were up 142% and our loss 18.6% down. Directors were satisfied with this trend and look forWARD to further gains in 2006/2007.

International business spearheaded from the MIDDLE East also grew. The Company supports customers from SE ASIA to North Africa and has an aggressive growth strategy encompassING Central and Northern Asia in 2006/2007.

Multiimedia's staff continue to work hard and professionally for all shareholders . Our structure is stable and strategy firm. The Board are to be also thanked for their diligence and attention to detail. I look forWARD to the New Year with enormous enthusiasm



### **Appendix 3 (a): MIT and SVT Instructions**

#### **Instructions**

The following task is a comprehension test based on the Chairman's Statement from a company's Annual Report. The test measures how easy the statement is to understand.

You are given a passage to read. Once you have finished reading, turn over the sheet of paper and read the test sentences. Each test sentence is to be marked as YES or NO. Answer YES if the test sentence has the same meaning as a sentence in the passage. Answer NO if there is no sentence in the passage with the same meaning as the test sentence.

The following short paragraph and test sentences illustrate the process:

The Trust will continue to seek opportunities to increase its assets and further diversify the agricultural risks of vineyard ownership. It will also pursue opportunities to acquire vineyards and other wine related assets at attractive prices with capital growth expectations. These acquisitions may lead to not only ownership but also operation of vineyards with associated agricultural risk. Other opportunities will also be sought involving other wine related assets, investments or joint ventures.

#### **Test questions.**

1. The Trust aims to spread the risk of vineyard ownership by increasing the number and type of assets held. **YES/NO**
2. It will also pursue opportunities to acquire vineyards and other wine related assets at attractive prices with capital growth expectations. **YES/NO**
3. Other opportunities will also be sought to divest wine related assets or investments. **YES/NO**

Test sentence 1 is marked YES since it paraphrases the first sentence of the paragraph. Test sentence 2 is marked YES as this simply restates the second sentence of the paragraph. Test sentence 3 is marked NO since there is no mention of divesting (selling off) assets or investments in the paragraph.

Your responses should be made from memory, based on your understanding of the passage. Do not turn back to the passage once you have started answering the questions.

## **Appendix 3 (b): MIT and SVT Reading Passage**

### **Chairman's Statement**

Your Company has achieved a great deal this year.

It concluded a hugely successful capital raising to purchase the intellectual property and Mediaport assets of NSS. Those assets are debt free and reside on 2 acres in Adelaide and Perth. The Mediaports support a world wide network of prestigious customers and boast long term profitable revenues. They are now the backbone of Multiimedia's future and will grow substantially as we acquire more customers, support more satellites and develop increasingly popular communication applications.

The Company sold Multie Technology Distribution Pty Ltd, its hardware distribution operation. This was in line with our strategy to divest of non-core operations.

The Board has received on 29 September 2006 a firm offer for the sale of Airworks Media Pty Ltd. This offer is under consideration and the Board expects to make an announcement shortly. A sale of Airworks Media Pty Ltd will be a wonderful achievement for a business founded in 2003 with just one employee.

Shareholders have voted to change Multiimedia's name to NewSat and in October this will take effect. The new name is well known in Australian telecommunications space and better reflects the Company's core business.

This year revenues for continuing divisions were up 142% and our loss 18.6% down. Directors were satisfied with this trend and look forward to further gains in 2006/2007.

International business spearheaded from the Middle East also grew. The Company supports customers from SE Asia to North Africa and has an aggressive growth strategy encompassing Central and Northern Asia in 2006/2007.

Multiimedia's staff continue to work hard and professionally for all shareholders. Our structure is stable and strategy firm. The Board are to be also thanked for their diligence and attention to detail. I look forward to the New Year with enormous enthusiasm.

## Appendix 3(c): MIT Questions

### Questions

1. Multiimedia acquired the media port assets of NSS. YES/NO
2. Those assets include unsettled debt and reside on 2 acres in Adelaide and Perth. YES/NO
3. The media ports purchased from NSS will provide revenues for years to come and support customers from all over the world. YES/NO
4. Multiimedia has purchased the local hardware distribution business, Multie Technology Distribution (MTD). YES/NO
5. Multie Technology Distribution was sold so the company can concentrate on its core business of satellite services. YES/NO
6. A firm offer to buy the subsidiary company Airworks Media Pty Ltd was accepted by the board on 29 September 2006. YES/NO
7. The Board has rejected an offer to buy Airworks Media and the outcome will be announced soon. YES/NO
8. The Company's core business is satellite communications and the new name of NewSat, which is already well known overseas, demonstrates this more effectively. YES/NO
9. Annual revenue for continuing divisions was 142% up from 2005 and the Company's total loss was 18.6% down from 2005. YES/NO
10. The directors anticipate further gains in 2006/07 following on from this years' satisfactory performance. YES/NO
11. 2005/06 also saw a growth in international business led by the Middle East. YES/NO
12. The Company is planning dramatic growth in 2006/07 and aims to achieve this by adding customers from Central and North America to its existing market in the SE Asia to North Africa region. YES/NO

## Appendix 3 (d): MIT Solutions

### Solutions

1. Multiimedia acquired the media port assets of NSS. {Paraphrase} YES/NO
2. Those assets include unsettled debt and reside on 2 acres in Adelaide and Perth. {Meaning change paraphrase} YES/NO
3. The media ports purchased from NSS will provide revenues for years to come and support customers from all over the world. {Paraphrase} YES/NO
4. Multiimedia has purchased the local hardware distribution business, Multie Technology Distribution (MTD). {Meaning change paraphrase} YES/NO
5. Multie Technology Distribution was sold so the company can concentrate on its core business of satellite services. {Paraphrase} YES/NO
6. A firm offer to buy the subsidiary company Airworks Media Pty Ltd was accepted by the board on 29 September 2006. {Meaning change paraphrase} YES/NO
7. The Board has rejected an offer to buy Airworks Media and the outcome will be announced soon. {Meaning change paraphrase} YES/NO
8. The Company's core business is satellite communications and the new name of NewSat, which is already well known overseas, demonstrates this more effectively. {Meaning change paraphrase} YES/NO
9. Annual revenue for continuing divisions was 142% up from 2005 and the Company's total loss was 18.6% down from 2005. {Paraphrase} YES/NO
10. The directors anticipate further gains in 2006/07 following on from this years' satisfactory performance. {Paraphrase} YES/NO
11. 2005/06 also saw a growth in international business led by the Middle East. {Paraphrase} YES/NO
12. The Company is planning dramatic growth in 2006/07 and aims to achieve this by adding customers from Central and North America to its existing market in the SE Asia to North Africa region. {Meaning change paraphrase} YES/NO

## **Appendix 4 (a): SVT Questions**

### **Questions**

1. Multiimedia acquired the media port assets of NSS. YES/NO
2. Those assets include unsettled debt and reside on 2 acres in Adelaide and Perth. YES/NO
3. In the past Multiimedia has had to lease media port services from other companies. YES/NO
4. The media ports purchased from NSS will provide revenues for years to come and support customers from all over the world. YES/NO
5. This acquisition will make the company one of the largest media port operators and satellite providers in the southern hemisphere. YES/NO
6. The Company sold Mutlie Technology Distribution Pty Ltd, its hardware distribution operation. YES/NO
7. The sale of Mutlie Technology Distribution was in line with our strategy to diversify operations. YES/NO
8. A firm offer to buy the subsidiary company Airworks Media Pty Ltd was received by the board on 29 September 2006. YES/NO
9. This offer is under consideration and the Board expects to make an announcement shortly. YES/NO
10. There were no dividends paid during the year and the Directors do not recommend the payment of a dividend. YES/NO
11. The Company's core business is satellite communications and the new name of NewSat, which is already well known in Australia, demonstrates this more effectively. YES/NO
12. This year revenues for continuing divisions were up 142% and our loss 18.6% down. YES/NO
13. Directors were satisfied with this trend and look forward to further profits in 2006/2007. YES/NO
14. International business spearheaded from Eastern Europe also grew. YES/NO
15. The decrease in cash inflow was mainly due to the acquisition of the NSS Australian business assets. YES/NO
16. The Company supports customers from SE Asia to North Africa and has an aggressive growth strategy encompassing Central and Northern Asia in 2006/2007. YES/NO

## **Appendix 4 (b) SVT Solution**

### **Solutions**

1. Multiimedia acquired the media port assets of NSS. {*Paraphrase*} YES/NO
2. Those assets include unsettled debt and reside on 2 acres in Adelaide and Perth.  
{*Meaning change*} YES/NO
3. In the past Multiimedia has had to lease media port services from other companies.  
{*Distracter*} YES/NO
4. The media ports purchased from NSS will provide revenues for years to come and support customers from all over the world. {*Paraphrase*} YES/NO
5. This acquisition will make the company one of the largest media port operators and satellite providers in the southern hemisphere. {*Distracter*} YES/NO
6. The Company sold Multie Technology Distribution Pty Ltd, its hardware distribution operation. {*Original*} YES/NO
7. The sale of Mutlie Technology Distribution was in line with our strategy to diversify operations. {*Meaning change*} YES/NO
8. A firm offer to buy the subsidiary company Airworks Media Pty Ltd was received by the board on 29 September 2006. {*Paraphrase*} YES/NO
9. This offer is under consideration and the Board expects to make an announcement shortly. {*Original*} YES/NO
10. There were no dividends paid during the year and the Directors do not recommend the payment of a dividend. {*Distracter*} YES/NO
11. The Company's core business is satellite communications and the new name of NewSat, which is already well known in Australia, demonstrates this more effectively.  
{*Paraphrase*} YES/NO
12. This year revenues for continuing divisions were up 142% and our loss 18.6% down.  
{*Original*} YES/NO
13. Directors were satisfied with this trend and look forward to further profits in 2006/2007.  
{*Meaning change*} YES/NO
14. International business spearheaded from Eastern Europe also grew.  
{*Meaning change*} YES/NO
15. The decrease in cash inflow was mainly due to the acquisition of the NSS Australian business assets. {*Distracter*} YES/NO

16. The Company supports customers from SE Asia to North Africa and has an aggressive growth strategy encompassing Central and Northern Asia in 2006/2007. {*Original*}  
YES/NO

