

METHODOLOGY

Open Access

An enduring lens for a continuing problem: Close analysis of conceptually scored *cloze* items

John M O'Toole^{1*}, Junyu Cheng² and Gjyn O'Toole¹

* Correspondence:

mitch.otoole@newcastle.edu.au

¹University of Newcastle, Newcastle, Australia

Full list of author information is available at the end of the article

Abstract

This paper provides an example of close analysis of student attempts to fill the gaps that comprise a *cloze* test. Such analysis may assist diagnosis of the difficulties that groups of students experience with specific text in particular contexts and illustrate the impact of those difficulties.

The paper briefly touches on the international, academic role of English, outlines its contemporary role in universities in the People's Republic of China and then describes a small quantitative study of student difficulty with specialist English that illustrates the impact of those difficulties on text accessibility.

One hundred and fourteen Chinese undergraduate students completed a 50 item *cloze* test which was conceptually scored before close analysis of patterns in clear student error.

The impact of the pattern of difficulty that emerged from the Chinese results is illustrated in a number of ways. The rank and proportional difficulty of the identified language features is presented in numerical form and compared with that emerging from previous use of the test in secondary school contexts. The proportional difficulty is then used to guide insertion of words from an unfamiliar language into the base text to provide an indication of the difficulty being experienced by the 'average' Chinese undergraduate in reading this passage prepared for a mid-level secondary school audience.

The present case provides the opportunity of 'proof of concept' for extension of *cloze* techniques from holistic estimation of readability to identification of specific features of specialist writing styles that may cause difficulty for particular groups of readers.

The results of this study suggest that such close analysis of error patterns may provide an illuminative lens on student difficulty with specific language styles within English and that, in this particular case, focus on general academic English may not be sufficient preparation for upper level discipline courses that make use of more specific styles.

Keywords: English for specific purposes; cloze tests; technical language; foreign language; curriculum

Background

Cloze testing has formed an enduring, if somewhat controversial, element of the language teaching repertoire for more than 60 years. Much of that controversy has revolved around use of the technique to provide an authentic window on text accessibility or test-taker language competence. This paper shifts focus from the whole

to the parts and describes a use of the technique to expose the specific difficulties that a particular group of students experience with specialist text. It represents an expansion of the conceptual scoring approach to interpretation of 'clozed' text. Close analysis of the clear errors that students make in attempts to fill gaps produced by the *cloze* procedure has the potential to predict the level of difficulty that students may have with other such text. Instructors often view the text they set as unproblematic and they may find the results of such analysis more convincing than general exhortation.

The *cloze* procedure has roots stretching back to attempts to quantify intelligence by Ebbinghaus and Minkus in the late nineteenth century, providing the somewhat odd label, derived from the German *closure*. The technique resurfaced in the mid-twentieth century (Taylor, 1953) before being popularised in the 1970s by Oller and his co-workers. The numbers provided by *cloze* techniques continue (Gellert & Elbro 2013) to be used to compare "the readability of texts, the language proficiency of students, the intelligibility of a given author" (Oller & Jonz 1994 p. 13).

Construction of *cloze* tests usually involves the deletion of a random word in the second sentence of a passage, then the deletion of every fifth (or seventh, ninth, eleventh or thirteenth) word thereafter until the desired number of words (usually 50) have been replaced by blanks (or a range of possible alternatives). A specific group of readers then tries to replace the words deleted from the particular passage. As Brown (2013) indicates, other variants of the technique include 'tailoring' of various sorts, so that test may involve deletion of words that produce the greatest degree of discrimination between students (as Brown himself did), deleting only content words (as can be seen in the 'activity' sections of many secondary school textbooks) or words of particular grammatical classes (as Tong et al 2014 did for Chinese conjunctions).

The words which readers suggest to replace deletions can be scored strictly (where only exact replacement of the word deleted will be coded as correct) or conceptually (where various synomic or otherwise meaningful alternatives might be acceptable). The group average of acceptable responses is interpreted as an estimate of the access the particular group would have to the meaning of the specific text.

However smooth the preceding account may appear, the controversy that emerged mid-century has never really faded. Spolsky (2000), referred to *cloze* as a "fad" (page 544) that required "selling" (page 545) in the face of doubts regarding its usefulness that were almost a decade old by the end of the 1960s. However, Stansfield (2008) went so far as to refer to the 1970s as the "decade of John Oller" (page 312). Nonetheless, there is contemporary disquiet about the naïve use of overall *cloze* scores (Brown 2013) and the methods that are used to defend their use (Sadeghi 2013).

Much of the controversy revolves around general interpretations of holistic results from raw, un-trialled tests that accumulate data from deletions representing items of widely differing difficulty. Conversely, *cloze* tests seem ideally suited to matching groups of *particular* students with *specific* styles of language, or to exploring the degree of mismatch. Patterns of student difficulty that emerge from close analysis of clear errors may guide instructor preparation and use of material that is supportive of student learning. A closer account of student difficulty with particular specialist styles may help institutions to respond more adequately to expanding student populations and to determine an appropriate balance between general and specific purpose language courses.

Universities are recruiting more international students and increasing numbers of people from outside the traditional local tertiary-oriented population have begun university study. Success for such 'non-traditional students' is far from assured in contexts that have been slow to adapt to this 'massification' (Fowler, 2008; Leathwood & O'Connell, 2003) and the role of wider tertiary literacy in relation to student success has emerged as a substantial issue of concern (Absalom & Golebiowski, 2002; Allan et al., 2013).

Parallel with this recent expansion, the rapid developments in science and technology occurring over the past 65 years have prompted more specific concern about how students read, understand and use science text in particular. Concern for the specific style of English used in Science is only part of the wider field of 'English for Specific Purposes' but the early recognition of the problems that scientific writing causes for learners makes it of enduring interest. Reviews indicate that the language of laboratory, educator, book and examination continues to thwart secondary school pupil and undergraduate student access to science, prompting teacher attempts to help those learners (O'Toole, 1996; Rollnick, 2000, Hand et al, 2010). Specialist staff can nonetheless remain unaware of the difficulties that their particular language style can cause for outsiders such as students.

China is an interesting example. It is a non-English speaking context, with a long tradition of independent scholarship, moving towards a position of great economic strength. The social and economic development of the People's Republic is diminishing historic disparities (Salager-Meyer 2008). This development has produced noticeable increases in student enrolment in Chinese universities. For example, the number of registered university students has risen from 6.4 million in 1995, through 23 million in 2005 to 27 million in 2007 (Zhou, 2009). China has a long history of examination-driven learning and contemporary testing causes some concern (Fan & Jin 2013).

Degrees in English for Specific Purposes have been offered by Chinese universities for a generation or more. Graduates from such courses often staff other foreign language departments. Undergraduates either major in English, or study the subject as part of another degree. The national curriculum for the latter suite of 'service' courses is called 'College English' (CIP 2007). Ten million students sat the College English Test (CET) in 2008 and almost 18 million in 2012 (see Jin 2014 for an insider perspective of the development of this examination) and its influence can only be described as "pervasive" (Li et al., 2012). The program leading to the CET normally occurs during the first two years of a four year undergraduate degree. There have been suggestions that extension for a further two years of more specialist language study could be useful.

College English provides more sustained and effective support for non-English majors than that which is available in many other contexts, including programs for foreign students in many English-speaking countries. Contemporary curriculum requirements essentially produce a general English program. Evaluation is considered to be a key to successful implementation and the summative phase usually takes the form of a pen and paper language examination; the CET is also available for web delivery but, at the time of writing, most students continue to take the paper-based test. It is therefore unsurprising that English courses in universities within China include rapid reading, intensive reading and extensive reading.

However, the difficulties that most students experience in their *technical* reading are not thoroughly dealt with in their College English courses. Consequently, it is also not

surprising that there should be a widespread perception that university students, most graduates, scientific specialists, engineers and other academic workers all confront great difficulties in their technical reading which prevent them from effectively accessing specialist information in their work and study. Such perceptions have been quantified in the distinctive Hong Kong SAR context (Evans & Green, 2007) as well as in contexts beyond China (Mazdayasna & Tahririan, 2008).

This interaction between a contemporary controversy in an interesting context and depth of work from other places prompted the present investigation of the utility of fine grained analysis of Chinese student attempts to replace deletions within a conventional *cloze* test. The over-arching research question is:

Does detailed analysis of the entries that undergraduate students make into a *cloze* test based on a mid-secondary school passage expose the areas of difficulty that they might experience in their wider and more specialist reading?

This investigation may function as 'proof of concept' for the extension of analysis of *cloze* results from the holistic to the particular. This may be an attractive proposition because of the potential the *cloze* technique holds for the direct use of authentic text with particular groups of students who are expected to read that text as part of access to, remediation of and extension on the content of their discipline courses.

This paper builds upon an earlier investigation of secondary school student language difficulties O'Toole & O'Toole (2004) The broad comparison of scientific reading performance between Chinese undergraduates who are EFL learners and other groups who speak English as their native or second language may provide insight into the challenges posed by reading scientific/technical material in English for both groups.

Beyond any 'proof of concept' function, the actual results in this case may be of interest for two reasons.

Firstly, undergraduates such as those participating in this investigation receive two years of College English and then move into the second half of their technical degree. This progression assumes that they will be able to read independently within their specialist area. The data from this study challenge that assumption.

Secondly, the performance of this group of students should be indicative of the language competence of Chinese-speaking students who may later attempt higher degrees that will require greater competence within specific styles of English. Such courses in Chinese universities and in English-medium universities both require students to read, understand and use technical text at an advanced level, regardless of the problematic nature of those expectations (Stokes & Martin 2008). Language has been identified as a challenge when enculturating Chinese academics into western institutions (Jiang et al., 2010) and it is reasonable to assume that it would be even more challenging for students. Although language is not the only source of difficulties in adjusting to English-speaking contexts (Zhou et al., 2008), there seems little doubt that it contributes to that difficulty.

Chinese undergraduate students have relatively limited access to English-medium contexts, suggesting that an existing test designed for younger, secondary school students

might be appropriate for the present investigation. The clear difficulty of the *cloze* task suggests that close analysis may over-estimate reader language difficulties but the difference between the students in this study and those for whom the test was first prepared suggests that this would be unlikely to exaggerate difficulties in the present case. These undergraduate students are required to read texts that are potentially more challenging than that forming the basis of the secondary test. The features identified by the present investigation may indicate difficulties that are rarely recognized and assist lecturers and material developers to predict possible areas of student confusion with the specialist text they encounter when they progress beyond the general style that is the focus for College English.

Methods

The Chinese undergraduate sample

In the present study, 114 Chinese sophomore students (19-20 years old) completed an existing test instrument (Additional file 1: Appendix 1, found in the separate Appendix file). These participants came from three different classes within various program majors. The individual major is less important in this investigation of the impact of a particular method of language analysis than it would have been if the primary focus was on the actual difficulty of these particular groups. Mandarin Chinese (or, more correctly, *Putonghua*) is the medium of instruction for most courses within the undergraduate student programs, although textbooks written in English may be used. The 114 students participating in this study were completing their second year of university study and, consequently, close to the end of their formal English instruction.

The instrument

The present investigation exposes language difficulties through the use of a fifth word deletion *cloze* test based on a 320 word sample that dealt with human defences against infection. The base passage was drawn from pages 211 and 213 of a secondary school science textbook (Heffernan & Learmonth, 2001). The book was written for students in Year 9 (13-14 years old). The particular passage selected as the basis of the *cloze* test had a Flesch-Kincaid (F/K) readability grade of 11, which suggests that while it may have been challenging for its initial audience it should be accessible for Chinese sophomore undergraduates (F/K 'grade 14').

The 50 item test previously demonstrated a reliability of 0.94 (Cronbach's α) in use with a sample of 806 junior secondary school pupils (ages ranging from 12 to 16 years) from schools in Australia, Singapore, the Philippines and Great Britain. The data from this earlier use of the test was re-analysed for comparison with that of the present investigation. The 806 secondary students were divided between those who claimed a monolingual English-speaking background (127 students) and those acknowledging more diverse language backgrounds (English as an Additional Language or Dialect, 'EALD': 679 secondary school students). This large latter group renders comparison of the test results more valid than it might otherwise have been. Most of the secondary EALD group are studying in contexts where English is only part of the local language repertoire. The health-based topic of the passage might reasonably be expected to form part of the general knowledge of young adults and its intended audience of 13-14 year olds makes its use with 20-21 year old undergraduate students defensible.

Deletion categorization

Each word deleted from the passage was labelled once by its traditional, or dictionary, classification and then again by a more modern grammar descriptor. These groupings allowed the formation of language feature sub-tests and discussion of separate results for individual language features.

The dictionary classifications (Crystal 2000) should require little explanation for the present audience; however the more modern descriptors may be more problematic. There is no established consensus (or perhaps, more correctly, there are several extant positions) regarding description of the structure of English. The more recent grammar descriptors used in this study lean heavily on earlier work by Michael Halliday (for example, Halliday 1990), although neither the earlier application of the test nor the present investigation made use of complete lexico-grammatical interpretation (such as might be provided by application of Gerot and Wignell 1994).

In this paper, *technicality* (Herbert, 1965; Martin, 1993) refers to the use of words which are either specific to the field ('fully technical') or used with particular meanings within that field ('semi-technical'). Technical (or semi-technical) words are often nouns but technicality is not restricted to words of that class. *Word stacks* (Stevens, 1977; Trimble, 1985) are a well demonstrated feature of scientific writing that consist of strings of adjectives, or other items functioning as adjectives, preceding an eventual noun whose meaning they progressively modify. Word stacks are also sometimes called nominal groups; adjective stacks or piles; pre-modifiers; and noun phrases. *Passive voice* (Cooray 1965, Kess, 1993; Trimble, 1985) refers to verbs and is an impersonal characteristic of scientific writing. *Cohesive devices* (Connor & Johns, 1990; Halliday & Hasan, 1967) are words or phrases that tie discourse together, signaling links in meaning across text. More detail on deletion classification is available in O'Toole & Laugesen (2011).

The two-stage categorisation was developed as a vehicle for encouraging pedagogical change in 'mainstream' classes (O'Toole 1998). For any such change to occur, non-language trained staff must be able to identify examples of the features that they expect their students to access. This is more probable if the categories are familiar, or at least used in readily available resources. The retention of traditional categories in readily available dictionaries is a strong argument for use of the traditional grammar classifications in this context.

Nevertheless, it remains true that dictionaries are word-based and relatively unconcerned with features above the sentence level. More modern grammar categories can compensate for this shortcoming in explaining the impact of characteristic dictionary categories, thereby providing a framework that is sensitive enough to expose the difficulties being experienced by different groups of students, powerful enough to explain why such difficulties may exist and accessible enough to allow the possibility of the analysis being used by change agents.

Deletion analysis

The student replacements of the words deleted to form the test preserved as Additional file 1: Appendix 1 were analysed in detail. Deletions which were filled with exactly the same word as was in the original passage were coded as instances of 'exact replacement'. Student suggestions of words which clearly fail to maintain the general meaning of the

passage were coded as 'error replacement'. A response was coded as 'conceptually correct' if it differed from the 'exact' term but its use maintained meaning in the passage. The broader conceptual scoring was based upon entries made by the most adept of the secondary school test takers in the earlier study. The marking guide produced for the earlier study was used in the present investigation to allow comparison of data emerging from school and university contexts (Additional file 1: Appendix 2, found in the separate Appendix file). To maintain consistency in conceptual scoring, all student scripts in the present investigation were marked by a single individual. Summing the exact and conceptually correct replacements yielded the 'conceptually correct total', while the sum of the clear mistakes yielded the 'error total'.

The low mean scores for *cloze* that are interpreted as ability to read text independently (eg., only 53% of deletions exactly replaced: Oller & Jonz 1994, p.6) suggest that readers may have difficulty completing *cloze* tests. Consequently, both the notion of 'conceptually correct' to be applied and the nature of the text on which the *cloze* test is based are crucial to this study. If the base text is too difficult or if the conceptual filter applied is too fine, close analysis of student entry errors may over-estimate and/or mis-identify student difficulty.

The notion of 'conceptually correct' used in this investigation is wider than that used in other studies (such as Kobayashi, 2002). The alternatives may not appear appropriate when the replacements are considered in isolation. For example, the wide range of alternatives accepted for item 36 (including *any*, *some* and *the*) all carry different meanings and would hardly be considered as synonyms. However, if the context of the passage as a whole is considered then it becomes clear that different students are reconstructing different passages while maintaining a semblance of the original meaning. This approach to conceptually correct coding recognises student response to the passage beyond the context of the words immediately surrounding the deletion. Comparison of Additional file 1: Appendices 1 and 2 (found in the separate Appendix file) would allow readers to reconstruct the range of various meanings students were building from the mutilated passage.

The use of a simple technical passage prepared for a much younger audience in combination with the application of a broader notion of correctness suggest that undergraduate difficulties may be under-recognised in this study, rather than over-stated.

If a student cannot suggest a deletion entry that maintains some meaning for part of a passage, it seems valid to assume that they have difficulty in understanding the deleted word. If they have difficulty with other words of the same type, the inference that a particular language feature is causing difficulty seems even more compelling. The error totals on the language feature sub-tests give an indication of student difficulty with particular features of the language of the test passage.

The test instrument seems fit for its limited purpose in this study and this methodology seems to avoid many of the objections to the *cloze* procedure raised in the literature.

Numerical analysis

Analysis was restricted to language feature sub-tests with reliabilities greater than 0.5 (Cronbach's α) in the previous study. The sub-tests for the dictionary categories *noun* ($\alpha = 0.77$), *article* ($\alpha = 0.59$), *verb* ($\alpha = 0.81$) and *preposition* ($\alpha = 0.52$) previously yielded defensible information as did the sub-tests for the more recent grammar descriptors

technicality ($\alpha = 0.69$), *word stacks* ($\alpha = 0.50$), *passive voice* ($\alpha = 0.60$) and *cohesive devices* ($\alpha = 0.75$). Table 1, contains both the sub-test means of the 114 Chinese undergraduates and some data emerging from previous work. The overall difficulty column rests on results from the test as a whole (50 items) while the sub-test columns yield a total of 40 dictionary category items and 36 more modern grammar category items. This is a consequence of discard of data from sub-tests that yielded inadequate reliabilities.

A multivariate analysis of variance (MANOVA) was used to compare the patterns of difficulty experienced by these 114 Chinese undergraduates with that suggested for 127 younger monolingual English-speaking pupils and 679 pupils who indicated that they used English as an additional language or dialect (806 secondary school students in all). These sample sizes (920 cases in total) ensure that the number of cases in each analysis cell will be substantially greater than the practical minimum of 30 and "MANOVA is fairly robust against violations of normality and equality of variance" under such conditions (Allen & Bennett 2010 p. 146).

The boxplot, histogram and Q-Q plots supported this expectation in the face of significant responses to both Kolmogorov-Smirnov and Shapiro-Wilk tests of normality for the overall error distribution and for the language feature sub-tests. Preliminary regression analyses revealed relatively high tolerances for all language features within each set of categories (from 0.412 to 0.626 for dictionary categories and from 0.497 to 0.665 for more modern grammar categories), indicating that multicollinearity would not impede interpretation and partial regression plots indicated linear relationships between the language feature categories. As underlying assumptions were sufficiently met, separate MANOVA analyses were undertaken with the test as a whole, items categorised into dictionary (noun, article, verb and preposition) and more modern grammar categories (technicality, word stacks, passive voice and cohesive devices) entered as the dependent variable and student sample groups as independent factors.

Multivariate analysis of significance suggested that apparent differences, within and between the student sample groups, are robust enough to permit discussion: overall ($df = 2$; $F = 3.2$; $p = 0.41$; $adjR^2 = 0.005$) and for both dictionary ($df = 2$; $F = 12.08$; $errdf = 1828$; $p = 0.000$) and more modern grammar ($df = 2$; $F = 41.99$; $errdf = 1828$; $p = 0.000$) item categories, with significance being equal under all tests but reported on the basis of Wilk's Lambda. Univariate comparisons were also significant ($p < 0.01$).

Results and discussion

The Chinese results on the *cloze* test appear on Table 1, which also includes secondary school data from the same test, marked on the basis of the same protocol. The three groups providing data for this table have substantially different backgrounds. The Chinese sample are undergraduates drawn from a single university, while the more numerous secondary samples are drawn from a number of schools in four nations. This may raise questions concerning the generalisability of the results. However, the re-use of existing instruments with varying populations is common research practice. Furthermore, it is well to recall the research question that guides this investigation. The 'proof of concept' function of this study makes recognition and

Table 1 Mean Sample Difficulty on English Language Features

Student Sample	No. of pupils	Noun % <i>wrong</i> n=18	At'cle % <i>wrong</i> n=5	Verb % <i>wrong</i> n=12	Prp'n % <i>wrong</i> n=5	Tchlty % <i>wrong</i> n=10	Word Stacks % <i>wrong</i> n=6	Pas'Ve Voice % <i>wrong</i> n=3	Cohsv Device % <i>wrong</i> n=15	Overall D'fctly % <i>wrong</i> n=50
U/G Chinese	114	53.17 (SD=13.73)	39.30 (SD=21.27)	51.90 (SD=21.12)	47.19 (SD=18.65)	47.81 (SD=20.51)	32.02 (SD=19.97)	42.98 (SD=28.65)	37.60(SD=17.87)	48.63 (SD=15.09)
	<i>Rank</i> [@]	1	4	2	3	1	4	2	3	
Secondary EALD[#]										
	679	42.89 (SD=18.06)	28.19 (SD=26.93)	39.80 (SD=25.84)	48.22 (SD=23.80)	31.69 (SD=20.29)	18.14 (SD=18.63)	36.38 (SD=33.19)	46.55 (SD=19.03)	43.41 (SD=19.83)
Secondary English	127	47.55 (SD=23.18)	30.08 (SD=28.72)	40.16 (SD=29.55)	41.10 (SD=26.94)	39.29 (SD=26.49)	25.33 (SD=26.05)	34.12 (SD=34.75)	45.88 (SD=23.36)	44.38 (SD=24.34)
	<i>Rank</i>	1	4	3	2	2	4	3	1	

Notes:

[@]*Rank* = Rank within category, most difficult = 1.[#]EALD = English as an Additional Language or Dialect.Table 1 is based on analysis of student responses on a single *cloze* test.

*Student data was recoded so that a clear error = 1 and acceptable replacement = 0.

**Cloze* test deletions were classified by language category.

*Two SPSS routines were written ("dictionary categories" & "modern grammar categories").

These did the following:

1 count number of items representing a particular language category (eg., nouns) coded as clear error.

2 divide that number by the number of items representing that category (eg., nouns) deleted (in this case, 18) to give the mean category (in this case, noun) difficulty.

3 multiply that mean by 100 to yield a percentage.

4 average that mean across the particular student sample (114 students for the undergraduate Chinese group: 53%).

The top row language descriptors are discussed previously, in **Deletion categorization**. The number at the base of each heading refers to the number of deletions categorized as reflecting that descriptor.

comparison of overall patterns of difficulty more important than contrast between specific language features in isolation.

Table 1 indicates that the three groups are all experiencing difficulty, although the pattern of difficulty varies somewhat. The difficulties assume practical importance when it is recognized that the results for all groups indicate the inability to provide broadly conceptually correct replacements for almost half of the deletions on this language test based on a passage from a mid-range secondary science textbook dealing with a topic of common interest. This suggests that any of these groups could experience difficulty in independently accessing such text, and further that they may have more trouble with resources that are more difficult to read.

It is evident that the Chinese undergraduate group experienced a greater degree of difficulty with more language features than either of the other two groups. The extent of secondary student difficulty may be somewhat surprising but the greater degree of difficulty of the undergraduate group is more so. The overall difficulties for the student groups on Table 1 (extreme right hand column) are calculated from the error total score data. That is, Chinese undergraduates averaged 49% (24/50) clearly **wrong** on this *cloze* test. The 'average undergraduate' was unable to provide conceptually correct replacements for close to half of the words deleted from this passage from a mid-range secondary school science textbook. It appears that these Chinese undergraduates, contrary to expectation, might have more difficulty with specialist language than either group of their secondary counterparts. This echoes results derived from different methodologies in other contexts (such as Ward, 2009).

The 'Rank' figures that appear in each cell indicate the relative difficulty of the indicated language feature for that sample of students. The fact that Table 1 suggests that the Undergraduate Chinese Sample found Technicality the most challenging of the more modern grammar categories (closely followed by the Passive Voice) is interesting. However, the fact that the Secondary English as an Additional Language or Dialect Sample had almost as much difficulty is of less interest than the fact that the EALD group found it to be the most difficult of that set of language features (as did their monolingual classmates). The *Rank* rows of Table 1 are likely to be of greater practical significance than comparison between the average error scores themselves.

Table 1 illustrates the potential that close analysis of student error in completing *cloze* tasks holds for recognising the patterns of difficulty being experienced by particular groups of students in reading specific text. Such recognition allows identification of specific problems and the development of particular ways of dealing with them.

The following passage should assist readers to appreciate the impact of the level of difficulty for student comprehension of discipline-specific texts, which the College English Requirements (CIP 2007) suggest should be achievable for all those emerging from Chinese undergraduate courses. The version below adjusts the base text passage according to the pattern of Chinese difficulty exposed in Table 1. Half of the verbs, prepositions and nouns and one third of the articles deleted from the base passage to form the *cloze* test have been replaced by words drawn from an Irish folk song, simulating the difficulty which the 'average' of the participating undergraduates would encounter in trying to read this text.

Natural defences against infectious diseases

When disease-causing microbes try to invade our bodies, we have a number of natural defences, many of which will fight any infection.

1. *Skin, mucus and 'bhaile*

When unbroken, the skin on the outside of the body and the moist layers on the skin that line the mouth, nose and lungs díolta keep out unwanted microbes. Hairs to keep out disease-carrying particles are contained in the nose, and any that get past the hairs into the nose or lungs are fógairt by the sticky mucus, and where tiny hairs, provided that they have not been killed by tobacco and marijuana smoke, remove mucus and trapped léanmhar quite efficiently. Bacteria that tsamhraidh the stomach are finally killed by its high acidity, although their spores can often pass through to the intestine. Gcreach in the tears can kill bacteria that try to invade the eyes.

2. *White ngéibhinn cells*

If the disease-causing léanmhar are able to get into the body through a burn or a wound, then the second line of defense bhúí into play, whereby any invaders are engulfed by some dhiaidh of scavenging white blood cells. Certain smaller white blood cells arrive in the blood and attack the invaders directly. Any leftover microbes and dead and dying cells are removed when a larger cell type dhiaidh later. Some invaders, such as worms, are too large anois white blood cells to engulf, so there is a third group of scavengers that seilibh enzymes outside their cell body, which then attack the bhreá of the invading parasites.

3. *Inflammation and fevers*

If the invaders overcome this attack by the antibodies and white blood cells, they will consequently start to multiply. The body responds by sending in more white ngéibhinn cells and antibodies and at the same time it also tries to block off the area invaded.

It is worth recalling that the data on which this text mutilation took place are based on mean group difficulties. Those students whose performance falls below that mean will experience considerably more difficulty (O'Toole & King 2011). Such lower-performing students are often the legitimate focus of instructor attention.

Detailed analysis of student attempts to 'cloze' the passage has yielded specific suggestions of particular difficulties. These results indicate that student difficulty extends beyond technical vocabulary. The fact that these Chinese undergraduates were unable to correctly replace almost half of the words deleted from the passage is a source of particular concern because their final years of undergraduate study will involve the expectation that they access tertiary textbooks written for native-speaking students. Those texts are considerably more difficult to read than the passage forming the basis of this cloze test.

Conclusions

Detailed analysis of conceptually scored cloze items does appear to allow fine-grained analysis of areas of student difficulty for potentially useful purposes.

Reference to the ranks presented on Table 1 indicates that these undergraduates had predictable difficulties with nouns and associated technicality, compounded by 'stacking' words into complex noun phrases. They had similar difficulty with verbs and the passive structures that are characteristic of such science text. This science passage should have been relatively easy for them to access. Prepositions and articles serve to identify and locate information in science text and the fact that these students appear to be missing those connections, as well as the explicit cohesive markers that were present in the passage, on more than one third of their appearances further compounds the slightly greater level of difficulty described above. Inability to meet the deliberately loose standards of 'conceptual match' used in this investigation, in more than a third of the cases taken to represent those features, casts doubt on the ability of these students to access this resource with the ease that might have been expected. Close analysis of student attempts at *cloze* allows detailed description of student difficulty.

Such detailed analysis also allows comparison with other student groups. While the undergraduate students in this investigation had more difficulty with this passage than the linguistically diverse secondary group (EALD), the pattern of difficulty was not identical. The secondary EALD group found prepositions particularly challenging, which was less the case for the Chinese undergraduates, for whom prepositions were the third most challenging feature, rather than the first. This pattern was repeated for text cohesion. This may well be due to the greater amount of direct language instruction involved in the College English program compared to mainstream science classes in English-medium contexts. These substantive comparisons are of some interest and they could guide the preparation of focussed language material for specialist contexts.

There is the broader question of who might be interested in knowing that a particular group of 114 Chinese undergraduates were unable to replace an average of 39.30% of articles (2 from 5 deletions) or 51.90% of verbs (6 from 12 deletions) with a conceptually correct entry?

The key to this is the word 'particular'. This detailed analysis technique allows those responsible for student learning at any level to select a passage that could appropriately represent the type of text from which they expect their students to learn. They can then generate a *cloze* test from that passage. This allows instructors to determine the particular features of the specialist style represented by the passage that are causing difficulty for students within their particular group. In this context, the differing difficulty of the individual deletions is pivotal, rather than being a challenge to text reliability (as noted by Brown 2013).

If it is possible to partition the group, it will also be possible to determine 'who' is having difficulty with 'what': 'who' being dependent on the sample sub-groups of interest and 'what' on the model of language chosen. The specific features of the specialist style that are isolated by this detailed analysis can be used to generate focussed language activities that could support 'Directed Activities Related to Text' (DARTs: UoC 2013). Attention to the relative levels of difficulty being experienced by the particular group as a whole makes it possible to illustrate the impact of these difficulties on student understanding. This can be persuasive for specialist teachers who assume that the language style used within their subjects is transparent to

students and that learner difficulties arise only from conceptual difficulties or lack of application.

Detailed analysis of the entries that undergraduate students make into a *cloze* test based on a mid-secondary school passage exposes specific areas of difficulty that they might experience in their wider and more specialist reading. Such analysis indicates that their difficulties extend beyond predictable problems with technicality. Their difficulties appear greater than those of younger students in English-medium contexts.

More significantly, this investigation suggests that patterns of student error in replacing words deleted to form *cloze* tests may fruitfully reward detailed analysis.

Additional file

Additional file 1: Appendices 1 and 2.

Competing interest

The authors declare that they have no competing interest.

Authors' contributions

All authors read and approved the final manuscript.

Author details

¹University of Newcastle, Newcastle, Australia. ²Southeast University, Nanjing, Jiangsu, China.

Received: 17 December 2014 Accepted: 20 April 2015

Published online: 16 May 2015

References

- Absalom, D, & Golebiowski, Z. (2002). Tertiary literacy on the cusp. *Australian Review of Applied Linguistics*, 25(2), 5–17.
- Allan, HT, O'Driscoll, M, Simpson, V, & Shawe, J. (2013). Teachers' views of using e-learning for non-traditional students in higher education across three disciplines (nursing, chemistry and management) at a time of massification and increase diversity in higher education. *Nurse Education Today*, 33, 1068–1073.
- Allen, P, & Bennett, K. (2010). *PASW statistics by SPSS: A practical guide, version 18.0*. South Melbourne, VIC: Cengage Learning.
- Brown, JD. (2013). My twenty-five years of cloze testing research: So what? *International Journal of Language Studies*, 7(1), 1–32.
- CIP. (2007). *College English curriculum requirements*. Beijing: Foreign Language Teaching and Research Press.
- Connor, U, & Johns, A. M. (Eds.). (1990). *Coherence in writing: Research and pedagogical perspectives*. Alexandria VI: Teachers of English to Speakers of Other Languages.
- Cooray, M. (1965). The English passive voice. *English Language Teaching*, 21(3), 203–210.
- Crystal, D. (2000). *Rediscover grammar*. Harlow: Pearson Education.
- Evans, S, & Green, C. (2007). Why EAP is necessary: a survey of Hong Kong tertiary students. *Journal of English for Academic Purposes*, 6(1), 3–17.
- Fan, J.S. & Jin, Y. (2013). A survey of English language testing practice in China: the case of six examination boards. *Language Testing in Asia*, 3:7 doi:10.1186/2229-0443-3-7
- Fowler, Z. (2008). Negotiating the textuality of further education: issues of agency and participation. *Oxford Review of Education*, 34(4), 425–441.
- Gellert, AS, & Elbro, C. (2013). Cloze tests may be quick, but are they dirty? Development and preliminary validation of a cloze test of reading comprehension. *Journal of Psychoeducational Assessment*, 31(1), 16–28.
- Gerot, L, & Wignell, P. (1994). *Making sense of functional grammar: An introductory workbook*. Cammeray: Antipodean Educational Enterprises.
- Halliday, MAK. (1990). Some grammatical problems in scientific English. *Australian Review of Applied Linguistics Series S*, 6, 13–37.
- Halliday, MAK, & Hasan, R. (1967). *Cohesion in English* (Vol. 9). London: Longman.
- Hand, B, Yore, LD, Jagger, S, & Prain, V. (2010). Connecting research in science literacy and classroom practice: a review of science teaching journals in Australia, the UK and the United States, 1998-2008. *Studies in Science Education*, 46(1), 45–68.
- Heffernan, D, & Learmonth, MS. (2001). *The world of science Book 3* (3rd ed.). South Melbourne: Pearson.
- Herbert, AJ. (1965). *The structure of technical English*. London: Longman.
- Jiang, X, Napoli, RD, Borg, M, Maunder, R, Fry, H, & Walsh, E. (2010). Becoming and being an academic: the perspectives of Chinese staff in two research-intensive UK universities. *Studies in Higher Education*, 35(2), 155–170.
- Jin, Y. (2014). The limits of language tests and language testing: challenges and opportunities facing the College English Test, in D. Coniam (Ed). *English Language Education and Assessment: Recent developments in Hong Kong and the Chinese Mainland*. (pp. 155-169). doi:10.1007/978-981-287-071-1_10
- Kess, JF. (1993). *Psycholinguistics: Psychology, linguistics and the study of natural language*. Amsterdam: John Benjamins.
- Kobayashi, M. (2002). Cloze tests revisited: exploring item characteristics with special attention to scoring methods. *The Modern Language Journal*, 86(4), 571–585.

- Leathwood, C., & O'Connell, P. (2003). 'It's a struggle': the construction of the 'new student' in higher education. *Journal of Education Policy*, 18(6), 597–615.
- Li, HL, Zhong, Q, & Suen, HK. (2012). Students' perceptions of the impact of the College English Test. *Language Testing in Asia*, 2, 77–94. doi:10.1186/2229-0443-2-3-77.
- Martin, JR. (1993). Technicality and abstraction: Language for the creation of specialised texts. In MAK Halliday & JR Martin (Eds.), *Writing science: Literacy and discursive power* (pp. 203–220). London: The Falmer Press.
- Mazdayasna, G, & Tahirian, MH. (2008). Developing a profile of the ESP needs of Iranian students: the case of students of nursing and midwifery. *Journal of English for Academic Purposes*, 7(4), 277–289.
- Oller, JW, & Jonz, J. (1994). *Cloze and coherence*. London: Bucknell University Press.
- O'Toole, JM. (1996). Science, schools, children and books: exploring the classroom interface between science and language. *Studies in Science Education*, 28, 113–143.
- O'Toole, JM. (1998). Climbing the fence around science ideas. *Australian Science Teachers Journal*, 44(4), 51–56.
- O'Toole, JM, & Laugesen, R. (2011). *Developing specialist language styles: Research and application*. Sydney: Boraga Academic.
- O'Toole, JM, & King, RAR. (2011). The deceptive mean: Conceptual scoring of cloze entries differentially advantages more able readers. *Language Testing*, 28(1), 127–144.
- O'Toole, JM, & O'Toole, G. (2004). Specialist language style and supposedly adept monolingual science students. *International Journal of Learning*, 9(2002/2004), 9, 1201–1214.
- Rollnick, M. (2000). Current issues and perspectives on second language learning of science. *Studies in Science Education*, 35, 93–122.
- Sadeghi, K. (2013). Doubts on the validity of correlation as a validation tool in second language testing research: the case of cloze testing. *Language Testing in Asia*, 3:15. doi:10.1186/2229-0443-3-15.
- Salager-Meyer, F. (2008). Scientific publishing in developing countries: challenges for the future. *Journal of English for Academic Purposes*, 7(2), 121–132.
- Spolsky, B. (2000). Language testing in 'The Modern Language Journal'. *The Modern Language Journal*, 84(4), 536–552.
- Stansfield, CW. (2008). Where we have been and where we should go. *Language Testing*, 25(3), 311–326.
- Stokes, P, & Martin, L. (2008). Reading lists: a study of tutor and student perceptions, expectations and realities. *Studies in Higher Education*, 33(2), 113–125.
- Stevens, P. (1977). Special purpose language learning: A perspective. *Language Teaching and Linguistics: Abstracts*, 10(3), 145–163.
- Taylor, W. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 30, 415–431.
- Tong, XH, Tong, XL, Shu, H, Chan, SF, & McBride-Chang, C. (2014). Discourse-level reading comprehension in Chinese children: What is the role of syntactic awareness? *Journal of Research in Reading*, 37(S1), S48–S70.
- Trimble, L. (1985). *English for science and technology: A discourse approach*. Cambridge: Cambridge University Press.
- UoC (2013). *Directed Activities Related to Text (DARTs)/Document* Cambridge University. [http://oer.educ.cam.ac.uk/wiki/Directed_Activities_Related_to_Text_\(DARTs\)/Document](http://oer.educ.cam.ac.uk/wiki/Directed_Activities_Related_to_Text_(DARTs)/Document) Accessed 20 October 2014.
- Ward, J. (2009). EAP reading and lexis for Thai engineering undergraduates. *Journal of English for Academic Purposes*, 8(4), 294–301.
- Zhou, Y. (2009). On teaching approach reform. *China University Teaching*, 1, 4–6.
- Zhou, Y, Jindal-Snape, D, Topping, K, & Todman, J. (2008). Theoretical models of culture shock and adaptation in international students in higher education. *Studies in Higher Education*, 33(1), 63–75.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
