PROPOSAL

Matching Second Language Readers to Texts in an ESP Sciences Course

(Submitted by Richard Noone for the Best Research Proposal Competition)

Abstract

This study investigates the strength of syntactic complexity and vocabulary difficulty to predict reading comprehension under various textual conditions using 6 different texts. It seeks to determine the relative importance of these two variables as well as resolve some controversy over which readings in FPEL 560 Science, Agriculture, and Engineering are more difficult for students to read. The findings of this study will shed light on issues related to the development of reading materials used for both testing and curricular purposes. In particular, the findings could result in specifications that allow testers and curriculum developers to devise texts that more accurately and reliably target particular levels of reading ability. For testing, this will make tests more valid and reliable. For curriculum developers, this will increase the effectiveness of instruction of what is arguably the most import skill that foundation students need to master.

Text Complexity Issues at SQU

At SQU currently, text complexity for both testing and curriculum is measured predominantly by the Flesch-Kincaid Grade Level readability formula (FKG) FKG. FKG is determined by a regression equation based on the average number of words in a sentence (to measure syntactic complexity) and the average number of syllables per words (to measure vocabulary difficulty). Using FKG, curriculum and testing have specified which texts are appropriate for the different levels of the programs. For instance, the Curriculum Document 2012-2013 states that Level 5 learners are expected to be able to handle texts with an FKG of 8-9, while Level 6 learners are expected to tackle texts from FKG 10-12 (Curriculum Unit, 2013). Moreover, the fact that FKG is only a guide and that other variables, including lexical density are a part of text complexity, is also mentioned in the document. There is no mention, however, of how import something like lexical density might be in comparison to FKG, and thus no guidance for a situation in which one must decide whether a text with more difficult vocabulary and a lower FKG is more or less difficult than a text with less difficult vocabulary and a higher FKG. This type of situation is exactly what recent research at the LC has uncovered for two sets of texts being used for level 5-6 FPEL Sciences, Agriculture, and Engineering students ([**ANONYMIZED NAME**], 2013) .

The texts in question are the units of in-house reading materials for use in the levels. In this case, one book has unit texts that were adapted to fit within the FKG requirements of Level 6, but which also underwent extensive vocabulary simplification ([**ANONYMIZED NAME**], 2013). The level of vocabulary simplification, moreover, was determined by experience working with students and running texts where students identified unknown words through a vocabulary profiler. The goal was to create materials that approximated the instructional level which [**ANONYMIZED NAME**] (2013) defined as texts with anywhere from 90 – 97 % known words. Subsequent research on three of the four texts created found that 76%, 67 %, and 88 % of the students were

reading at the instructional level respectively ([**ANONYMIZED NAME**], 2013). Given the variation in student ability present within the level, this indicates a successful attempt at instructional match using these criteria.

As the book written by [**ANONYMIZED NAME**] was intended for Level 6, it came after the Level 5 book in the curriculum. However, [**ANONYMIZED NAME**] found that his students had persistent questions regarding large numbers of unknown words from the Level 5 texts and both he and his students felt that they were significantly more difficult that the texts designed for Level 6. This prompted [**ANONYMIZED NAME**] to investigate and compare the complexity of the Level 5 and Level 6 books. His investigation found that although the Level 5 book texts had consistently lower FKG scores, the vocabulary complexity in terms of word frequency was significantly greater ([**ANONYMIZED NAME**] 2013). More specifically, vocabulary from the Level 5 texts had a significantly smaller percentage of words coming from the first 1000 of the General Service List (GSL), and texts from the Level 6 book had significantly longer sentences which resulted in higher FKG. In his report, [**ANONYMIZED NAME**] (2013) concluded that the texts with the easier vocabulary (higher percentage of words from the first 1000 of the GSL – Level 6 texts) were the easier texts for three reasons. First, readability formulas are notorious for being misused and, consequently unreliable when used to simplify texts, especially when sentences are shortened in order to lower readability measures as had been done in the case of the lower FKG texts for Level 5 (Blau, 1982). Secondly, the primary predictor variable of most readability formulas –including FKG- is the vocabulary measure (Chall, 1981) and so changes made to a text that alter the measure of the formula but do not take vocabulary into account clearly invalidate the measure. Thirdly, student feedback indicated the high vocabulary texts were more difficult to read. As a result of this conclusion, the texts for the Level 5 Sciences book now come after the texts originally intended for Level 6 in the FPEL Sciences Curriculum so that students encounter progressively more difficult texts as they go through the course.

Nevertheless, controversy remains, and the texts in Agriculture and Engineering remain in their original order with those written for Level 5, but with more difficult vocabulary, coming before those written for Level 6, but with significantly easier vocabulary. One possible reason for this is that sentence length is a good measure of syntactic complexity (Chall, 1981) and some educators believe syntactic complexity to be more important to reading comprehension than vocabulary and so a text with shorter sentences will be less syntactically complex and, therefore, easier to read. Research that has investigated the relative contribution of vocabulary knowledge and syntactic knowledge to reading comprehension generally supports the idea that vocabulary is more important (Tweissi, 1998), but there is research that finds the opposite (Shiotsu & Weir, 2007), and this indicates there is still uncertainty as to the relative importance of these variables.

Research Questions

1. Are there any differences in student's ability to comprehend the texts written for Level 6 with low vocabulary complexity and high FKG and the texts written for Level 5 with high vocabulary complexity and low FKG?
2. If so, what are they?

3. What are the relative contributions of syntax as measured by sentence length and vocabulary complexity as measured by word frequency to reading comprehension and how do these variables interact?

Study Design

As mentioned above, [**ANONYMIZED NAME**] (2013) found the percentage of words from the first 1000 of the (GSL), and syntactic complexity as measured by sentence length to be the two primary variables that distinguish the Level 5 from the Level 6 texts. In this case, the Level 6 texts have a higher percentage of words from the first 1000 of the GSL resulting in low vocabulary complexity, but have longer sentences resulting in high FKG measures. On the other hand, the Level 5 texts have a much lower percentage of words from the first 1000 of the GSL resulting in high vocabulary complexity, but have shorter sentences resulting in low FKG measures. Thus, in order to thoroughly investigate the roles that these two variables play (percentage of words from the first 1000 of the GSL, and changes in FKG that result from changes in sentence length), three texts from each book (Level 6, and Level 5) have been selected and adapted to reflect all possible combinations of these two variables. Different groups will then be given these texts and their reading comprehension will be measured and compared using two different tests –a multiple choice reading comprehension tests with item types similar to our current EOLTs at the LC, and a fixed ratio cloze test. There will also be a topic familiarity measure given to control for any effects this variable may have had.

The two methods chosen for measuring of reading comprehension were chosen for a number of reasons. First of all, as multiple choice comprehension tests have been used in so much previous research into instructional match, and text variables that affect reading comprehension (Blau, 1982; Shiotsu & Weir, 2007; Tweissi, 1998) it makes good sense to utilize them here to facilitate the comparability of findings. However, these kinds of tests provide one limitation for this kind of research which could conceivably threaten the validity of its findings. This limitation is that, even with similar specifications, it is possible to produce tests that are more or less difficult, not as a result of the texts, but as a result of the items (Alderson, 2005). Moreover, it is difficult to ensure that multiple choice test items are capturing the differences between the complexities of texts because it is only possible to design reliable items on some of the information in a text and the information tested may be more or less difficult to comprehend than the text is overall. One solution to these problems is using random deletion cloze tests since these tests evenly sample all parts of a text and do not require item writing. Therefore, reading comprehension will be measured using random deletion cloze tests in conjunction with standard multiple choice comprehension tests.

Instruments

Piloting
Prior to using these instruments in the study, they will be piloted with students in the Fall semester to ensure that items are working, revise or add items, confirm time needed, and ensure reliability.

Texts

The texts used in this study have been carefully selected to ensure both their comparability and adaptability. One issue that has limited the comparability of different texts is the role that visual aids play. Because of the role of visual aids, it was only possible to realistically compare certain texts. For example, Cell Towers is always compared to Electricity-Wind because both have visuals often referred to in the text that require complicated decoding. Another issue was overall coherence. At least two texts were rejected because of their lack of general coherence which did not allow for changes in variables without extensive revision. Text adaptations were made using the average measures of the existing texts from both books using both FKG and the % of words from the 1st 1000 of the GSL (List 1). These two variables differ systematically across the two books with the average FKG of Level 5 being 8.6 and the average percentage of List 1 words being 75.5, while the average FKG of Level 6 is 10.5 and the average percentage of List 1 words is 80.7.

Multiple Choice Reading Comprehension Tests (MCRCTs)

One MCRCT was designed for all text conditions of each text. For example, all students who read any version of the DNA text will take the same MCRCT. Questions on these tests were selected so that they were able to be answered by all text conditions. Additionally, as items that use exact words from texts or have similar vocabulary to that used in the text are easier (Buck, Tatsuoka, & Kostin, 1997) care was taken to ensure that all language used in the items did not present an advantage or disadvantage to any text condition to the extent that it was possible to do so. Items on the MCRCTs tested objectives reflected by existing reading tests in the EFP here at the LC. Item formats include a majority of four option multiple choice questions, but also some two option True or False items. From 24 to 36 items were constructed for each text.

Cloze Tests

Research into cloze tests for L2 reading assessment has indicated numerous variables that can make such tests easier or harder regardless of the actual difficulty of the passage (Abraham & Chapelle, 1992). The most influential of these variables for fixed-ratio cloze tests, however, is the ratio of content versus function words as function words are found to constitute easier items (Abraham & Chapelle, 1992). In this study, the purpose of the cloze tests is to create a reading comprehension measure that captures the actual difficulty of the texts. Therefore, content/function word ratios on these tests were determined by analyzing what those ratios are on the texts from which they were created and all cloze tests are within 1 percentage point of those ratios. Because texts of about 1000 words are too long for a full cloze test, two versions of each text with from 92 to 112 items each were created. Between these two versions, almost all of each text is covered and at least one paragraph is kept the same between each version for purposes of test linkage via subsequent Rasch analysis which will nullify the results of any sampling bias.

General Reading Comprehension Measures

Reading scores from the previous EOLT and the midterm test will be used as a general reading comprehension measure to ensure comparability of groups and to measure any reading growth that may have occurred with the different groups used in the study.

Topic Familiarity Measure

Research into L2 reading comprehension has found that topic familiarity can play a role in reading comprehension processes (Pulido, 2007). Therefore, a measure of topic familiarity will be given

to students in the form of a question (How much did you know about the information in this text before you read it?) measured via a 5 point Likert scale (almost all, a lot, some, very little, almost nothing). Questions will also be asked to confirm if any students studied or reviewed information relevant to the readings between the first and second reading class (students participating in the study will be asked not to do this).

Procedure

Once approval from the LC administration and support of the relevant CCs has been achieved, the next step is to identify and recruit teachers who are willing to participate in the study. It is hoped that 12 or more sections distributed between Sciences, Agriculture, and Engineering will participate. Once the participating sections have been identified, the researcher will look at the schedules and assign different conditions as outlined in the Study Design to different sections until all conditions have been met. In order not to disrupt the normal flow of curricular content, the texts will be administered to groups at the time that those texts actually appear in their pacing schedule. One result of this is that different groups will end up doing the study at different times.

For all teachers, materials for each of the two reading classes will be given at least one day prior to the first class. These materials will include instructions and all measurement instruments. Teachers will distribute the materials according to the procedure below:

Reading Class 1:
1. Briefly explain the purpose of the study and give out consent forms (they give more details about the study) for students to sign. (10 minutes)
2. Distribute MCRCTs (1 hour)

Reading Class 2:
1. Distribute the Cloze Tests (half A and half B) (1 hour)
2. Distribute the Topic Familiarity Measure (10 minutes)

The researcher will collect administered instruments for Reading Class 1 & 2 from teachers after Reading Class 2 is finished and begin marking of the MCRCTs tests right away. Because the responses will be recorded on OMR marking sheets, they need not be scored by teachers or the researcher. Instead the researcher will run them through the scanner for marking within two days of receiving them. The researcher will then give the results of the MCRCTs to the teachers so that they can go over the students' results (if they wish). The researcher will then engage in data analysis.

Data analysis and ethical procedures

Ethics
All students with whom the study will be conducted will have its purpose and benefits both for themselves and LC explained to them. They will then be given handouts to be signed which indicate their willingness to participate. Anyone may choose not to participate. As a total of four reading classes will be utilized in conducting the study, the value of how that time is spent by students has been carefully considered. In most cases, the text read is already a part of the curriculum and the teacher has the option of using the study time to cover that text or letting it supplement existing materials. The first day of each data collection period the students will be given a reading comprehension test whose format is just like EOLTs in the LC. This kind of practice is something that students generally want more of. Additionally, the second class will give

students another test format for which they can compare their abilities. Finally, to make up for lost class time, classes involved in the study may opt not to spend class time using existing practice tests. When all these factors are considered, it would be hard to argue that participation in the study disadvantages students and it may that other sections wish to participate.

Data Analysis

A large number of different techniques will be applied to the data collected in this study. The use of Rasch measurement in constructing interval scales for the different variables to be compared is an advance over previous research designed to investigate these questions. Rasch measurement techniques will also be used to ensure the comparability of different groups, link cloze tests, equate MCRCTs, and equate general reading comprehension measures in order to measure reading growth. Moreover, to analyze interactions between different text conditions (independent variables) topic familiarity, MCRCTs, and cloze tests (dependent variables) MANCOVA will be used. To explore how well the variables of syntax and vocabulary difficulty as measured in this study are at predicting reading comprehension, both hierarchical multiple regression and structural equation modeling will be compared.

Resources required for the conduct of research

More than anything, this study requires the support of the LC administration, CCs, teachers, and students. Use of the latest version of SPSS is also needed for data analysis as well as the updated version of Winsteps for Rasch analysis. Moreover, several boxes of paper will be required. Of course any release time offered would be of great assistance as well.

References

Abraham, R. G. & Chapelle, C. A. (1992). The meaning of cloze test scores: An item
    difficulty perspective. *The Modern Language Journal*, 76(4), 468-479.
Alderson, J. C. (2005). *Assessing Reading*, Cambridge, UK.: Cambridge University Press.
Blau, E. K. (1982). The effect of syntax on readability for ESL students in Puerto Rico.
    TESOL Quarterly, 16(4), 517-528.
Buck, G., Tatsuoka, K., & Kostin, I. (1997). The subskills of reading: Rule-space analysis of a
    multiple-choice test of second language reading comprehension. *Language Learning,*
    47(3), 423-466.
Chall, J. S. (1981). *Readability: Conceptions and misconceptions*. Urbana, Illinois: National
    Council of Teachers of English.
Curriculum Unit (2013). *Foundation Programme English Language Curriculum Document*
    2012-2013, Muscat: Sultan Qaboos University.
[**ANONYMIZED NAME**],**. (2013).
Pulido, D. (2007) The effects of topic familiarity and passage sight vocabulary on L2 lexical
    inferencing and retention through reading. *Applied Linguistics,* 28(1), 66-86.
Shiotsu, T. & Weir, C. J. (2007). The relative significance of syntactic knowledge and
    vocabulary breadth in the prediction of reading comprehension test performance.
    Language Testing, 24(1), 99-128.
Tweissi, A. I. (1998). The effects of the amount and type of simplification on foreign language
    reading comprehension. *Reading in a Foreign Language,* 11(2), 191-206.