Using content language analysis system indexes (CLASI) in the development of English testing materials

Using Content Language Analysis System Indexes (CLASI) in the Development of English Testing Materials

Steven L. Renshaw

Abstract

CLASI consists of readability measures, grammar based indices, and word list metrics related to second language learning and general text analysis. Attention has been given to using CLASI as a base for subjectively determining appropriate passages for use in language testing, and some development of indices for use in judging materials has been developed. However, specific quantitative research has not been reported.

This article presents research related specifically to the use of CLASI in determining appropriate reading materials for an English language testing situation. Results indicate that several measures are correlated with average comprehension scores. The multiple regression model of significantly (bivariate) correlated variables (with the exception of the Flesch index) on comprehension was not significant. However, the regression model using the Flesch Index itself was significant, and this index appears to be the best general gauge of reading difficulty, at least for the testing situation studied. Qualifications of results as well as prospective applications in testing situations are discussed.

Content Language Analysis System Indexes (CLASI) was developed by Renshaw (2003, 2004, and 2006) as a computer based text analysis tool for use in research and pedagogy of second language learning, primarily in the context of Japanese learners of English. CLASI indices have been shown to be correlated with judged proficiency of English usage. Developments and usage of the program have

been discussed at length.

Attention has been given to using CLASI as a base for selecting testing materials (specifically texts to be used in reading, oral production, and listening examinations), and Renshaw (2006) discusses three indices (RDI, ORDI, LDI) that are a part of CLASI (see Table 2):

(a) Reading comprehension (Reading Difficulty Index, or RDI base) in which students read a passage and then respond to a number of multiple-choice questions.

(b) Reading production (Oral Reading Difficulty Index, or ORDI base) in which students are judged on their ability to read a given passage fluently and comprehensively, and

(c) Listening ability (Listening Difficulty Index, or LDI base) in which student listen to a given passage and then respond to a series of multiple choice questions.

There are at least two major difficulties in developing such indices: (1) Samples for statistical analysis must consist of the readings or texts themselves, rather than subjects who respond, and some average measure of response to the reading must be used as a data point. Sample sizes are by their very nature small and based on specific administrations. (2) Since measures must be derived from responses subjects make to items developed by the test administrators, the quality of the items themselves has a direct effect on the scores derived and may or may not validly reflect text difficulty. Given these considerations, the three indices mentioned by Renshaw have heretofore been included in CLASI analysis for heuristic purposes only, and caution has been indicated in using them in any actual situation.

After several years of test administration, a collection of 22 readings used in a

yearly reading comprehension examination at KUIS has been built. This corpus, while still small, provides a somewhat larger base than previously available for assessing specific correlations between CLASI readability indices and average scores of comprehension following reading of an English passage. Given these considerations, the fundamental question of the research is whether or not CLASI analysis can be used to determine probable difficulty of readings as measured by judgments of subject response to comprehension questions following the reading of a passage. The situation is somewhat different from any of the three mentioned indices already incorporated into CLASI but most closely related to ORDI.

Following description of the basic indices of CLASI, procedures and results of analysis are presented. This is followed by a general discussion of results and implications for use of CLASI in testing situations.

CLASI Indices

A substantial description of the initial indices of CLASI is presented in Renshaw (2003, 2006), and the set of initial and subsequent indices are shown in Tables 1 and 2* . The Academic Word List (AWL, see Coxhead, 2000) and sub lists have now been added. The fundamental numerical basis of the measure (count, index,

* Requirements for running CLASI include a Pentium based machine with 32 MB memory (minimum recommended), Windows OS 95 or higher (Windows 2000 or above recommended), and 16 bit color capability with at least 1024X768 display resolution. Hard disk requirements are minimal and depend on the amount of data under consideration. Benchmark estimates for processing 500 messages of 100-150 word length in batch mode are less then two minutes on a Pentium II based system with 128 MB memory running Windows 2000. CLASI is available for download from the following URL: http://www2.gol.com/users/stever/clasi.htm. After download, unzip the files to a temporary directory and run the SETUP program. Visual Basic source code is available from the author upon request.

or index based grade level), method of determination (formula, dictionary lookup, simple count) and literature source are provided in the tables. Note that RDI, ORDI, and LDI are included with the above caveats.

Development of a Corpus of Readings and Comprehension Scores

The corpus of readings used in this research came from successive administrations of an examination conducted at KUIS each year. This exam is used to determine admission to the university, and results are held in strict confidence. Since only correlations of scores with CLASI indices were used in this research, not specific scores and outcomes, there was no danger of divulging private information. However, for student confidentiality and university security, only correlations and results of regressions are reported here. Specific content of readings and score averages are not provided. It should be noted that this examination is a specific procedure developed for a specific university, and results should be viewed in that light. While testing procedures may be questioned, the purpose of this research is not to criticize specific procedures but rather to determine the possibilities for using CLASI as an aid in selecting text corpus, whatever the testing procedure of a specific administration or university.

Readings used in the examination process are developed by a university committee. The committee procures readings from a variety of resources. Readings are discussed and edited in committee with an attempt to make them as homogeneous in difficulty as possible. In the examination itself, a student reads the prepared passage and responds to questions about its content. Students are then given a comprehension score based on whether or not they accurately respond to these questions. Both a native speaker and Japanese instructor of

Table 1. Measures used in CLASI

INDEX	NUMERIC	ALGORITHM	SOURCE
English as Foreign Language (EFL)	Percentage of Total 0-100	Dictionary Lookup Based on British Natinal Corpus	Reynolds (2002)
Monbusho Japan High 1 (JHS1)	Percentage of Total 0-100	Dictionary Lookup EIKEN Level 1	Reynolds (2002)
Monbusho Japan High 2 (JHS2)	Percentage of Total 0-100	Dictionary Lookup EIKEN Level 2	Reynolds (2002)
Monbusho Japan High 3 (JHS3)	Percentage of Total 0-100	Dictionary Lookup EIKEN Level 3	Reynolds (2002)
Monbusho English Junior High School (MES JHS)	Percentage of Total 0-100	Dictionary Lookup Monbusho Recommendation for Junior High	Reynolds (2002)
Dale-Chall	Indexed to Grade	Formula	Dale & Chall (1948)
Flesch	Indexed to Grade	Formula Flesch Reading Ease	Flesch (1974)
Kincaid	Indexed	Formula Flesch/Kincaid Grade	Johnson (2004)
Spache	Indexed	Formula	Spache (1953)
Coleman-Liau	Grade Level	Formula	Johnson (2004)
Lix	Indexed	Formula	Björnsson. (1968)
Ari	Indexed	Formula	Smith & Taffler (1992)
Fog	Grade Level	Formula	Johnson (2004) Gunning (1952)
SMOG	Indexed	Formula	McLaughlin (1969)
Words/Sentence	Number of Words Divided by Total Number of Sentences	Counted Average	NA
Characters/Word	Number of Alphanumeric Characters Divided by Number of Words	Counted Average	NA
Syllables/Word	Vowel/Consonant Combinations Divided by Total Number of Words	Counted Average	NA
Misspelling/Words	Number of Misspelled Words Divided by Total Number of Words	Dictionary Lookup	NA
Words	Total	Counted	NA
Sentences	Total	Counted	NA
Paragraphs	Total	Counted	NA

Table 2. Measures Added to CLASI since 2004 Version

INDEX	NUMERIC	ALGORITHM	SOURCE
Reading Difficulty Index (RDI)	Indexed	Regression Based Formula	Author (see Text)
Oral Reading Difficulty Index (ORDI)	Indexed	Regression Based Formula	Author (see Text)
Listening Difficulty Index (LDI)	Indexed	Regression Based Formula	Author (see Text)
Kelk 1000 Common Words	Percentage of Total 0-100	Dictionary Lookup	Kelly & Kelly (2003)
Kelk T3000	Percentage of Total 0-100	Dictionary Lookup	Kelly & Kelly (2003)
5000 Collegiate Words	Percentage of Total 0-100	Dictionary Lookup	Kelly & Kelly (2003)
Ogden General Things	Percentage of Total 0-100	Dictionary Lookup	Ogden (1932a, 1932b, 1937, 1969)
Ogden Picturable	Percentage of Total 0-100	Dictionary Lookup	Ogden (1932a, 1932b, 1937, 1969)
Ogden Qualities	Percentage of Total 0-100	Dictionary Lookup	Ogden (1932a, 1932b, 1937, 1969)
Ogden General English	Percentage of Total 0-100	Dictionary Lookup	Ogden (1932a, 1932b, 1937, 1969)
Academic Word List and Sub Lists	Percentage of Total 0-100	Dictionary Lookup	Coxhead (2000)
Adverbial	Number of Adverbs Divided by Total Number of Words	Counted Average	NA
Prepositional Phrases	Number of Prepositions Divided by Number of Words	Counted Average	NA
Articles	Number of Articles Divided by Total Number of Words	Counted Average	NA

English evaluate responses. Since they are generally allowed to discuss their scoring, correlations between these judges are generally above 0.98. In most cases, 80 to 100 students respond to each reading, and pairs of judges (native speaker and Japanese) see from 6 to 10 subjects for a given reading.

For this research, an overall average score was calculated for all students responding to a particular reading. There has been little difference in administrative procedure relative to the judging process from year to year.

From 2003 through 2008, a total of 22 passages with resulting average comprehension scores were made available and selected for this research. The comprehension scores represent a measure of the difficulty of each passage, again given the caveats of such mentioned earlier.

CLASI Analysis and Data Preparation

All 22 readings were coded and run through CLASI in batch mode (see Renshaw, 2006). An XLS file was produced, and average comprehension scores for each reading were added. This file was then read into SPSS for analysis. Bivariate correlations were conducted between average comprehension scores and CLASI indices. Step-wise multiple regression was then conducted using significantly correlated ($p<.05$) CLASI indices as independent variables and average comprehension as dependent variable.

Results of Statistical Analysis

Significantly correlated CLASI indices ($p<.05$) are shown in Table 3. Indices are not only highly correlated with comprehension but also with one another. Most of these indices are based on sentence length, word length, and average syllables per word. Sentence length and word length are common to all tabled indices with the exception of Flesch. Flesch incorporates syllables per word and sentence length. While average syllables per word is significantly related to comprehension scores, sentence length and word length are not (statistically at least). It is interesting that these latter measures are incorporated into other significantly correlated indices (FOG, Kincaid, Dale/Chall). Both Flesch and Kincaid incorporate syllables per word, but formulas derived by their respective authors differ.

Table 3. Indices Significantly Correlated with Average Comprehension Scores

Correlation

		Compre-hension	Sylables per Word	Flesh Index	FOG Index	Kincaid Index	Data/ Chall
Comprehension	Pearson Correlation	1	.566 **	-.601 **	.601 **	.578 **	.475 *
	Sig. (2-tailed)		.006	.003	.003	.005	.026
	N	22	22	22	22	22	22
Sylables per Word	Pearson Correlation	.566 **	1	-.959 **	.756 **	.806 **	.584 **
	Sig/ (2-tailed)	.006		.000	.000	.000	.004
	N	22	22	22	22	22	22
Flesh Index	Pearson Correlation	-.601 **	-.959 **	1	-.853 **	-.941 **	-.679 **
	Sig. (2-tailed)	.003	.000		.000	.000	.001
	N	22	22	22	22	22	22
FOG Index	Pearson Correlation	.601 **	.756 **	-.853 **	1	.876 **	.784 **
	Sig. (2-tailed)	.003	.000	.000		.000	.000
	N	22	22	22	22	22	22
Kincaid Index	Pearson Correlation	.578 **	.806 **	-.941 **	.876 **	1	.716 **
	Sig. (2-tailed)	.005	.000	.000	.000		.000
	N	22	22	22	22	22	22
Dale/Chall	Pearson Correlation	.475 **	.584 **	-.679 **	.784 **	.716 **	1
	Sig. (2-tailed)	.026	.004	.001	.000	.000	
	N	22	22	22	22	22	22

**. Correlation is significant at the 0.01 level (2-tailed)
*. Correlation is significant at the 0.05 level (2-tailed)

Table 4 shows results of step-wise regression using indices reported in Table 3 as independent variables and average comprehension scores as dependent. While directional, the model is not significant ($p<.10$). Coefficients are reported in Table 5. None are statistically significant. From this data, an index based on these regressed CLASI indices appears unwarranted. Note that Flesch index is excluded as being redundant.

Table 4. Regression of Significantly Correlated Indices in Comprehension

ANOVAb

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	33.394	4	8.349	2.729	.064[a]
	Residual	52.001	17	3.059		
	Total	85.395	21			

a. Predictors: (Constant), Dale/Chall, Syllbles per Word, Kincaid Index, FOG Index
b. Dependent Variable: Comprehension

Table 5. Regression Coefficients for Indices

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	1.452	14.071		.103	.919
	Syllables per Word	8.475	11.634	.237	.729	.476
	FOG Index	.490	.634	.349	.773	.450
	Kincaid Index	.150	.915	.073	.164	.872
	Dale/Chall	.030	.896	.010	.033	.974

a. Dependent Variable: Comprehension

The regression model and coefficient for the Flesch index are reported respectively in tables 6 and 7.

Table 6. Regression Using Flesh Index Only

ANOVAb

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	30.821	1	30.821	11.295	.003[a]
	Residual	54.574	20	2.729		
	Total	85.395	21			

a. Predictors: (Constant), Flesch
b. Dependent Variable: Comprehension

Table 7 Regression Coefficient for Flesh Index

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	35.126	4.563		7.698	.000
	Flesh Index	-.228	0.68	-.601	-3.361	.003

b. Dependent Variable: Comprehension

Syllables per words is obviously a very strong indicator of difficulty as measured by average comprehension scores. However, the Flesch index which incorporates syllables per word with other measures in its formula appears to explain a larger amount of variation in comprehension scores.

Discussion and Prospects

There are obviously many points where error variance can enter this study. In some ways, it is based on averages of averages, and data points are not

individual measures of ability but rather average of abilities of many subjects. As mentioned earlier, the measures themselves are based on items created to test comprehension of texts, not inherent measures of texts themselves. Study passages were not randomly selected, and the limited sample size of 22 is in itself a source for concern in any interpretation of results.

Given the above qualifications, there are some interesting indications in this research. The Flesch index itself is based on regression using average syllables per words as well as sentence length. It was developed by Flesch (1974) with specific data at a specific time. The author has sometimes heard TESOL colleagues remark that "Flesch is best" when assessing text. Perhaps results of this research show this assertion to have a measure of truth. Results using the corpus of this study indicate that by itself, the Flesch index can account for higher levels of variation in comprehension scores than any other measure or combination of measures in CLASI. Thus, there appears to be no need to add yet another regression formula index to CLASI, at least when it comes to this type of reading in this type of testing situation. It will be interesting to see if similar results are obtained, not only with larger samples of similar readings in similar situations, but in other contexts where English texts are chosen and edited for language testing. Obviously, more rigorous study of RDI, ORI, and LDI with larger samples seems appropriate Hopefully, such research will be forthcoming.

Is CLASI necessary to calculate a Flesch index? Of course not. Many text analysis programs provide the Flesch index. However, this index may be used with Ogden measures, the AWL, Monbusho indices, and other word/phrase lists indices to select and assess appropriate readings. CLASI provides a central source of a number of measures that can be used in text evaluation. Difficulty is of course

not the only criteria for determining what is appropriate for a particular examination procedure. Obviously, the specific needs and criteria of individual testing situations determine what text indices may be most useful.

References

Björnsson, C. H. (1968) *Läsbarhet.* Lund: Liber.

Coxhead, A. (2000) A New Academic Word List. TESOL Quarterly, 34(2): 213-238.Dale, E. & Chall, J.S. (1948) A Formula for Predicting Readability, *Educational Research Bulletin*, 27, 221-233.

Flesch, R. (1974) *The art of readable writing*, New York: Harper and Row.

Gunning, R. (1952) *The technique of clear writing*. New York: McGraw-Hill.

Kelly, C. & Kelly, L. (2003) Brian Kelk's Top 1,000 UK English Words. http://www.manythings.org/vocabulary/lists/l/kelk-about.php. Accessed November 18, 2006.

McLaughlin, G. (1969), SMOG grading: A new readability formula. *Journal of Reading, 12* (8) 639-646.

Ogden, C. K. (1932a) *The Basic Dictionary, being the 7,500 most useful words with their equivalents in Basic English, for the use of translators, teachers and students*. London, K. Paul, Trench, Trubner & Co., Ltd.

Ogden, C. K. (1932b) *The Basic Words, a detailed account of their uses*. London, K. Paul, Trench, Trubner & Co. Ltd..

Ogden, C. K. (1937) *Basic English, a general introduction with rules and grammar*. London, K. Paul,Trench, Trubner & Co., Ltd.

Ogden, C. K. (1969) *Basic English: International Second Language*. New York, Harcourt,Brace & World.

Renshaw, S.L. (2003) Content Language Analysis System Indexes (CLASI): A

Computer Based Program for Text Assessment, *Studies in Linguistics and Language Teaching,* 14, 31-47.

Renshaw, S.L. (2004) Content Language Analysis System Indexes (CLASI) and the Prediction of Judged Writing Ability, *Studies in Linguistics and Language Teaching,* 15, 1-17.

Renshaw, S.L. (2006) Current Development and Usage of Content Language Analysis System Indexes (CLASI), *Studies in Linguistics and Language Teaching,* 17, 151-166.

Reynolds, B. (2002) English as a Second Language. http://eslsv001.esl.sakuragaoka.ac.jp/teachers/BR/. Accessed November, 2004

Smith, M & Taffler, R. (1992). Readability and understandability: Different measures of the textual complexity of accounting narrative. *Accounting, Auditing & Accountability Journal* (4) 84-98.

Spache, G. (1953) A new readability formula for primary grade reading materials. *Elementary School Journal.* (55) 410-413.