TOEFL, CULTURAL BIAS AND LATIN BACKGROUND

Heitor Garcia de CARVALHO, Maria Inês GARIGLIO & Sidney Carlos de MELO (Centro Federal de Educação Tecnológica de Minas Gerais - CEFET)

Abstract

Through a validity analysis of a TOEFL sample applied to Brazilian undergraduates, the linear relationship between their scores and silent reading of authentic, technical-scientific texts was rejected. Thus, its uset as co-variate in the statistical analysis has been discarded and a more complex model of variables has been postulated.

Key-words: *reading; TOEFL; cultural background; technical-scientific texts.*

Resumo

Através da análise de validade de amostra do TOEFL aplicada a graduandos brasileiros, a relação linear entre seus resultados e a leitura silenciosa de textos técnico-científicos autênticos foi rejeitada. Assim, seu emprego como co-variada na análise estatística foi descartado e um modelo complexo de variáveis foi postulado.

Palavras-chave: *leitura; TOEFL; background cultural; textos técnico-científicos.*

1.   Introduction

The Test of English as a Foreign Language - TOEFL - is widely used to select foreign students who apply for undergraduate and graduate studies in universities in English-

speaking countries (Hoesley & Meredith, 1979, apud Vollmer & Sang, 1983). But some of the findings of research on the variables involved in the processing of authentic, technical-scientific texts casted doubts on TOEFL adequacy, when applied to students whose cultural-linguistic background is Latin. One initial hypothesis assumed a linear relationship between word knowledge, grammatical command and silent textual processing (O'Donnell, 1962, apud Harris, 1971). Books used in teaching English as a second language, published by big publishing houses in English-speaking countries, seem to assume a similar hypothesis. However, after research this simple model had to be abandoned, a more complex model postulated, and the usage of the TOEFL for co-variance statistical analysis considered inadequate.

As the main research issue was the intervening variables on visual text processing, some literature was consulted. It related that the early studies dealing with reading ability and instruction have focused on word knowledge (Gray, 1925, apud Harris, 1971) and most of the experiments reported were conducted with disabled children (Blank and Bridger, 1966, apud Harris, 1971) or with those belonging to American social minorities (Justman, 1965, apud Harris, 1971; Labov, 1973, apud Cummings, 1983). According to Cummings,

> Language minority students are frequently transferred from bilingual to English-only classrooms when they have developed superficially fluent English communicative skills. Despite being classified as "English proficient" many such students may fall progressively further behind grade norms in the development of English academic skills (e.g., see Mazzone, 1980). Because these students are relatively

fluent in English, it appears that their poor academic performance can no longer be explained by their English language deficiency, and thus cognitive or cultural "deficiencies" are likely to be invoked as explanatory factors. (1983:110)

On the grounds of intellectual correlates of reading, the linguistic ability of reading was, at that time, linearly related to the intelligence quocient: the more words a child was able to recognise the more intelligent and greater was his/her reading ability (Hilliard, 1924; Bond and Clymer, 1955; Gray, 1960; Potts, 1960; apud Harris, 1971). And according to Cummings, "this assumption leads directly to the conclusion that poor performance on an L2 verbal IQ test is a function of deficient *cognitive* abilities (i.e., learning disability, retardation)". (1983: 110)

According to the view of the reading process as essentially a process of thinking in relation to the written symbol, of meaning elaboration, lists of most frequent words, like the one provided by Thorndike (1917, apud Harris, 1971), have been produced to reduce the difficulty of texts intended to teach reading and thus make the learning process easier.

The readability formulas were designed out of the need for greater precision, objectivity, and efficiency in assessing the reading difficulty of materials and, therefore, to help instructors to select the material to be read by their students. According to Dale and Chall (1956, apud Harris, 1971), the major components of the readability formulas were concerned with the employed vocabulary, the sentence structure, the density of the expressed idea, and the human interest in the selection to be read. Zipf (1935, apud Carrell,

1987) worked out a readability formula (the *Zipf Law*) related to word length: a long word is less frequent and, therefore, more difficult.

Most of these pieces of research have focused on the material to be read. The input, then, was external to the reader.

On the other hand, a shift of research perspectives was marked by Ausubel's theory (1963, apud Oliveira, 1973) of advanced organizers which pressed a more cognitive profile on the reading theories, since it took on account the reader itself. The reading process was not focused in word knowledge any more, but in the meaning the author, through text, tries to convey, and the meaning constructed by the reader. Due to that focus shift, learning to read and reading to learn could be considered as different actions related to textual processing: the first enhancing the command of basic reading abilities and the second, the meaning. Authors like Rumelhart (1980), Eskey (1987), Cohen (1990), Fairclough (1991), Lévy (1994), among others, see the reading process as an integrative activity, where the reading material is interpreted and meaning constructed according to the previous experiences the reader has already stored in the long term memory and hierarchly organized in his/her schemata.

Then, it was pertinent to examine the categories embedded in the learning process. Valette's (1971) cognitive behavior description is based upon Bloom's Taxonomy of Educational Objectives. So, the objectives are hierarchly listed from the simpler (knowledge and perception) to the more complex (understanding and production). According to her, the ability of word recognition is one of the simplest cognitive behaviors in language learning.

Prompted by these findings, an inspection of current teaching materials of English as a Second Language provided some ascertainings, focusing on the teaching material:

* Though working with different points of view, the lists of more frequent words (Thorndike, 1917 apud Harris,1971) and the Zipf Law (1935, apud Carrell, 1987) gave rise to a practice, which seems to be a methodological assumption, adopted in the teaching material design, of vocabulary control and hence the elimination of longer Latin-origin words.

And from other sources (Dale and Chall, 1956, apud Harris, 1971; Krashen, 1983):

* grammar structures grading: the simpler ones must be taught first;
* the text structure: the didatic texts must present situations taken from the students' generic world knowledge.

These principles related to vocabulary and grammar are also found in current, widely used software devoted to textual correction (e.g. the grammarchecker) as well as in the teachers' section of English teaching books. A cursory inspection of didatic texts shows that they deal with the students' common world knowledge (shopping, travelling, casual conversation, etc.), and present simple text structure.

Thus, it is correct to assert that research has historically been grounded on word knowledge (Shoben, 1980); so, most of the teaching material enhances the cognitive behavior of word recognition and the employment of more complex ones is sparse.

2.   Statistical validity analysis

There are many specific considerations in choosing a test, such as validity, reliability, and practicality. The foremost questions to be asked are "How valid is it?", "What do we want it to measure?". A test of reading achievement must measure comprehension, which comprises high-level skills on Bloom's Taxonomy. Both content and statistical validity should be taken into account. Our assumption that TOEFL sample presented a linear relationship with reading comprehension of authentic, technical-scientific texts was proved not to be sound. Then a broader content analysis was carried out.

First assumptions took for granted that the knowledge about grammar and vocabulary would be reflected on the textual processing (O'Donnell, 1962, apud Harris, 1971), so the hypothesis assumed that students who present a good performance on TOEFL vocabulary and grammar sub-tests would have similar results in the reading comprehension sub-test.

So, the broad research design provided three levels of contrast between: 1- Engineering undergraduates (with exposition to the instrumental approach) versus Human Sciences undergraduates (with no exposition to the instrumental approach); 2- groups with 30, 60, and 90 hours of the instrumental approach; 3- mechanical and electrical engineering undergraduates. Besides the TOEFL sample, five other tests, prepared from authentic, technical-scientific texts, were applied to the subjects. Since all these five tests measured reading ability, the subjects' perfomance in TOEFL was to be used as co-variate in the statistical analysis.

The TOEFL sample employed was drawn from three different versions of authentic TOEFL tests and was applied to 91 Brazilian undergraduates in the Industrial Engineering course of Centro Federal de Educação Tecnológica de Minas Gerais (Federal Centre of Technological Education of Minas Gerais) and to ten Human Sciences undergraduates of Universidade Federal de Minas Gerais (Federal University of Minas Gerais). Due to the original research design, the oral abilities were not tested, so the testees answered 20 multiple-choice questions on each of the following sub-tests: reading comprehension, vocabulary, and grammar. The students had one and a half hours to answer the 60 questions. As a consequence of the observed results of the pilot test, carried out to monitor possible problems, half of the students received a test in the following order: grammar-vocabulary-reading; the other half, reading-vocabulary-grammar, to avoid the stress on the last sub-test, which showed up in the pilot test. But, for the statistical treatment, the results were all converted to the same order: grammar-vocabulary-reading.

The statistical analysis with the Spearman-Brown formula (Guilford, 1956) gave a reliability index of 0.90. Such an index, despite modifications on the application procedures, shows that, on the whole, the instrument was appropriate. In the item analysis, the distribution of the difficulty index was roughly adequate but the validity indexes detected some problems, since 26 were classified with low validity index and 22 with very low validity index; and the discrimination indexes showed that 50 % of them were below 0.19.

Thus, even though the test was reliable and with a roughly adequate difficulty index, it showed low validity to discriminate students with best ability to read and interpret

authentic, technical-scientific texts. Besides that, the research question was proved to be optimistic and unreal: the students with the best performance for the grammar and vocabulary sub-tests did not present similar performance for the reading comprehension sub-test nor in the tests carried out in other stages of the original research.

Once the hypothesis of linear relationship between TOEFL scores and silent reading ability of authentic, technical-scientific texts proved to be inconsistent, its employment as co-variate in the statistical co-variance analysis proposed by the initial research design coud be discarded and a more complex model has been postulated, which is open to future research. However, to find out which factors were affecting these results and their behavior on the new proposed model, further TOEFL analysis has been carried out, focusing on content validity.

A thorough content analysis has pointed out several test construction features to be accounted for in the reduced TOEFL scores utility as co-variate in the initial research design. Some of them were due to inherent flaws on the assumptions of norm-based test construction. A normal distribution of abilities in the tested population is usually assumed and also that testees might constitute a random sample of this universe[1]. But, as results showed, the students whose cultural background is Latin configure a homogeneous

---

[1] According to Carroll, "the validity of a proficiency test was dependent not only on how representative the chosen sample of elements of English was but also on the degree of certainty with which the success of the examinee in solving future tasks and handling certain social situations could be predicted on the basis of his/her test results. (...) From the classical point of view of test theory it was argued that the degree of mastery in a foreign language could be inferred from the multitude of measured language performances (at least four). Thus it would be possible to form some sort of overall picture of a person's language ability and (by way of generalization) of his/her ability to act in a more or less predictable manner even in future situations requiring language use." (1972, apud Vollmer & Sang, 1983: 33-35)

sub-group with distinctive advantages and disadvantages in TOEFL. Item analysis has showed that students with high total scores have also reached higher scores on the vocabulary sub-test. But high scores on the vocabulary sub-test did not guarantee the same perfomance for the reading sub-test.

3.   Content validity analysis

A thorough analysis was carried out on the grammar sub-test. It did not show a high validity in relation to total scores. Probably because of norm-based test construction assumptions, the items must be independent from each other and randomically chosen, which lead to an apparent statistical lack of internal consistency. Thus, the grammar sub-test questions seemed to test items chosen from a grammatical universe as an itemized and discrete collection. Then, a high score in this sub-test has not meant a greater command in the English language and did not show relevance to textual processing. This has strongly negated the idea of linear dependency between reading ability and awareness of grammar and vocabulary formulated by O'Donnell (1962, apud Harris, 1971).

One of the distinctive advantages that students of Latin cultural background had had in this specific TOEFL sample resulted from the emphasis on word recognition and the massive contrast between Latin and non-Latin derived words. This contrast seems to be rooted in the Zipf Law, by which Latin words are difficult for native speakers but, in fact, easy for Latin students (see Table 1). On the vocabulary sub-test, all 20 questions proposed such contrast. This strategy tends to make those questions easier for those who, having Latin cultural background, are proficient in their native language. Therefore, the difficulty

and discrimination indexes have shown that the assumption that Latin-origin words are difficult turned out to be false for such a group: testees were reading the words in the proposed alternatives as if they were written in Portuguese, since most of them were almost perfect cognates. Some examples can be provided:

The reading comprehension sub-test content analysis showed that it was mainly a word recognition task, since it just required the constrast between a word taken from the text and the proposed alternatives[2].

Besides that, the employed texts presented common scripts belonging to scientific texts which are difficult for native children but far easier for undergraduate and graduate applicants, due to their technical content.

Table 1
Comparison between English and Portuguese word spelling

---

[2] One example taken from a reading sub-test question: it was written "annually" in the text, and in the correct option "once a year" was found

English Words	Portuguese Correspondents
proprietors	proprietários
tutors	tutores
benefectors	benfeitores
debtors	devedores (débito)
correct	correto
meticulous	meticuloso
incessant	incessante
accurate	acurado
ancient	antigo (ancião)
reside in	residir em
accommodate to	acomodar a
invade	invadir
abandon	abandonar

Source: TOEFL sample (Gariglio, 1996).

As it was previously said, the reading comprehension sub-test dealt heavily with the cognitive ability of word recognition which does not find ressonance on the daily routine of undergraduate and graduate students who are required to apply higher abilities.

4.   Discussion

Some TOEFL features, which may be causing problems, seem to be rooted in the same underlying assumptions of widely recommended methodologies for teaching English as a Foreing Language (EFL). Perhaps their assumptions might be embedded in the unruffled tradition from the early findings on teaching reading abilities to American children and seem to have been transferred to the didatic texts aimed at the adult international student: emphasis on word recognition; vocabulary selection; grammar grading; texts based on generic world knowledge. Most

of the English-teaching books examined, aimed at native children and adult foreign students, have shown an underlying tendency towards the elimination of Latin-origin words. The intention is to make texts easier for learners, which is in accordance with the Zipf Law (1935, apud Carrell, 1987) and in convergence with Thorndike's (1917, apud Harris, 1971) lists of most frequent words. On the other hand, for applicants for advanced studies, these same words, taken as difficult ones for native students, are the most widely used in technical-scientific texts. A comparison between authentic and didatic texts (see Table 2[3]) demonstrated that the former had 67% of Latin cognates and the latter only 7,5%.

Table 2
Differences of lexical items of didatic and authentic texts

	Texts	Authentic		Didactic	
	Total number of words	361		132	
Lexical Aspects	Frequency / %	F	%	F	%
	Cognates	67*	18.5	10	7.5
	Technical jargon	8	2.2	0	0
	Verbal nominalization	14	3.8	1	0.7
	Technical NP	6	1.6	0	0

*one false cognate
Source: Gariglio, 1996.

But, the abilities required from the applicants to fulfill their aims, especially in reading authentic, technical-scientific texts, are far different from those. These assumptions bring

---

[3] Table 2 is presented to illustrate the differences on lexical aspects between texts: one authentic, taken from a USA magazine on computers, and the other, a didatically prepared one, taken from a British English teaching text-book.

about some "side effects" for foreign students whose cultural background is Latin and for applicants for undergraduate and graduate studies. They are expected to accomplish reading tasks heavily based upon technical, scientific texts and are less required, for instance, to speak to native people or discuss domestic or mundane issues. This provides an interesting contrast: the discussion of these matters is easy for native speakers and may be more difficult for foreign students. According to Olson, "(...) [the] processing of text calls for comprehension and production strategies which are somewhat different from those employed in everyday speech and which may require sustained 'education' for their acquisition." (1977, apud Cummings, 1983:116). On the other hand, scientific subject matters may present some difficulty to native laymen and may be easier for foreign applicants for undergraduate and graduate studies. Taking for granted that what is difficult for native speakers is also difficult for foreigners, the TOEFL reading sub-test employs scientific texts, which turns out to be an artificial difficulty[4]. As an example, one of the texts, presented in the employed TOEFL sample, dealt with chemical reactions involved in photosynthesis. The testees, who were mature readers, were able to answer the related five questions due to their previous knowledge, since this matter is learnt in high school.

Besides that, the texts have not presented a fixed script like those belonging to the conventional behavior of tourists, shopping, foreing students, weddings etc., which constrains the didactic texts (usually employed in the applicants' English language learning process) and helps the

---

[4] By taking Bruner's contrast between communicative competence and analytic competence, it is possible to assert that foreign graduation candidates present mastery on the latter one, "since it involves the prolonged operation of thought processes exclusively on linguistic representations. It is made possible by the possession of communicative competence and is promoted largely through formal schooling." (1965, apud Cummings, 1983: 115)

students' generic world knowledge to contribute to make them far easier than those which do not have such a constrained scenario. The English teaching texts, for regular schooling, are "constructed" to meet the requirements imposed by a syllabus of structures to be followed and by a rigid gradation of vocabulary items. In addition, they are not authentic texts, do not present the genre profile which characterizes the authentic technical-scientific texts, nor the relative simplicity of vocabulary and extention, in number of words, like those employed on conventional text-books for English language teaching. Moreover, didactic texts usually serve as prompts for the listening and speaking abilities, since the conventional English teaching methodology is clearly based upon a hierarchical organization of the basic linguistic abilities: listen first, then speak, and after that, read and write.

These findings should prompt a revision of such methodological assumptions, suggesting a restructuring of didatic texts for the adult foreign student.

5.   Conclusions

In TOEFL, it seems that the testee's cultural and linguistic background does not play an important part in the construction of the test, which may interfere in the final results, since the emphasis on the constrast Latin/non-Latin words seems to make the test easier for those with Latin background and highly educated. Besides that, Saxon words may be easier for native children and frequent in everyday conversation of private matters but rare in authentic, technical-scientific texts and printed material.

As it has been said before, the cognitive behavior involved in the TOEFL vocabulary and reading

comprehension sub-tests accomplishment is based on word recognition. Besides that, the TOEFL structure seems to contradict both major aspects of the current English-teaching material: the usage of artificially difficult vocabulary (Zipt Law) and texts which do not present strict everyday scripts. So, it is possible to especulate that students who take the test are interested in graduate courses and, therefore, their future studies would require much more cognitive effort than just reading to recognise words. They are supposed to learn concepts, principles and problem solving strategies. Mere word recognition cognitive behavior probably would not guarantee the measurement of those cognitive abilities. Thus, as a proficiency test, those sub-tests seem to be far less demanding than the daily academic routine the foreign undergraduate students would face. Taking the reading comprehension sub-test as an example, it seems fairly clear that the required ability is the word recognition: the testees are not expected to make inferences, synthesis nor evaluation, but to perform a contrast between words explicitly written in the text and in the proposed answers.

References

AUSUBEL, D.P. (1963) The psychology of meaningful verbal learning. IN: Grune & Straton (eds.) *Introduction to school learning*.

BLANK, M. & W.H. BRIDGER (1966) Dificiencies in verbal labeling in retarded readers. *AM J Orthopsychiatry*, **36**: 840-47.

BOND, G.L. & T.W. CLYMER (1955) Inter-relationship of the SRA primary mental abilities, other mental characteristics, and reading ability. *J Ed Res*, **49**: 131-36.

BRUNER, J.S. (1975) Language as an instrument of thought. IN: A .Davies (ed.) *Problems of language and learning*. London: Heinemann.

CARRELL, P. (1987) Content and formal schemata in ESL reading. *TESOL Quaterly*, **21**.3: 461-81.

CARROLL, J.B. (1972) Stalking the wayward factors: review of Guilford and Hoepfner's. The analysis of intelligence. *Contemporary Psychology*, **17**: 321-24.

COHEN, A.D. (1990) *Language learning and problems of knowledge*. Cambridge: The MIT Press.

CUMMINGS, J. (1983) Language proficiency and academic achievement. IN: J. W. Oller (ed.) *Issues in language testing research*. Rowley: Newbury House Publishers, Inc.

DALE, E. & J.S. CHALL (1956) Developing readable materials. IN: *Adult reading*. 55th Yearbook, Part II, NSSE, U. Chicago.

ESKEY, D.E. et al. (1987) *Teaching second language for reading for academic purposes.* s.l.: s.m.

FAIRCLOUGH, N. (1991) *Language and power*. Essex: Longman.

GARIGLIO, M. I. (1996) *Leitura de textos autênticos em inglês em contexto de ensino de ciência e tecnologia.* Dissertação de Mestrado. Belo Horizonte: Centro Federal de Educação Tecnológica de Minas Gerais.

GRAY, W.S. (1925) *Summary of investigations relating to reading*. Chicago: University of Chicago Press.

_____ (1960) Reading. IN: C.H. Harris (ed.) *Encyclopedia of educational research*. 3rd ed. New York: Macmillan.

GUILFORD, J. P. (1956) *Fundamental statistics in psychology and education.* 14th edition. Tokyo: Kógakusha & McGraw Hill.

HARRIS, T. L. (1971) Reading. IN: C. H. Harris (ed.) *Encyclopedia of educational research.* New York: Macmillan.

HILLARD, G.H. (1924) *Probable types of difficulties underlying low scores in comprehension tests*. State U. Iowa.

HOSLEY, D. & K. MEREDITH (1979) Inter- and intra-test correlates of the TOEFL. *TESOL Quarterly*, **13**: 209-217.

JUSTMAN, J. (1965) Academic aptitude and reading test scores of disadvantaged children showing varying degrees of mobility. *J Ed Meas*, **2**: 151-55.

KRASHEN, S. D. (1983) The input hypothesis. IN: J. W. Oller (ed.) *Issues in language testing research*. Rowley: Newbury House Publishers, Inc.

LABOV, W. (1973) The logic of nonstandard English. IN: N. Keddie (ed.) Tinker, Tailor, *The myth of cultural deprivation*. Harmondsworth: Penguin.

LÉVY, P. (1994) *As tecnologias da inteligência - o futuro do pensamento na era da informática.* Rio de Janeiro: Editora 34.

O'DONNELL, R. (1962) Awareness of grammatical structure and reading comprehension. *H Sch J*, **45**: 184-88.

OLIVEIRA, J.B. (1973) *Tecnologia educacional - teorias da instrução*. Petrópolis: Vozes.

OLSON, D.R. (1977) From utterance to text: the bias of language in speech and writing. *Harvard Educational Review*, **47**: 257-28.

POTTS, E. (1960) A factorial study of the relationship between the child's vocabulary and his reading progress in the infant's stage. *Br J Ed Psychol*, **30**: 84-6.

RICH, E. (1983) *Artificial intelligence*. New York: McGraw Hill Book Company.

RUMELHART, D. E. (1980) Schemata: the building blocks of cognition. IN: Spiro et al. (ed.) *Theoretical issues in reading comprehension - perspectives from cognitive psychology, linguistics, artificial intelligence and education*. Hillsdale: Lawrence Erlbaum Associates.

SHOBEN, E. J. (1980) Theories of semantic memory: approaches to knowledge and sentence comprehension. IN: Spiro et al. (eds.) *Theoretical issues in reading comprehension - perspectives from cognitive psychology, linguistics, artificial intelligence and education*. Hillsdale: Lawrence Erlbaum Associates.

THORNDIKE, E.L. (1917) Reading and reasoning: a study of mistakes in paragraph reading. *J Ed Psychol,* **8**: 323-32.

VALLETE, R.M. (1971) Evaluation of learning in a second language. IN: B. Bloom et al. *Handbook on formative and summative evaluation of learning*. New York: McGraw-Hill.

VOLLMER, H. J. & F. SANG (1983) Competing hypothesis about second language ability: a plea for caution. IN: J. W. Oller (ed.) *Issues in language testing research*. Rowley: Newbury House Publishers, Inc.

ZIPF, G.K. (1935) *The psycho-biology of language*. Boston: Houghton-Mifflin.

Heitor Garcia de Carvalho, PhD in Educational Technology - Concordia University, Canada. Associate Professor - Research and Post Graduation Department of CEFET-MG. Main areas of interest: reading; cognition; computers and education; artificial intelligence; hypertext.

Maria Inês Gariglio, MSc. in Technology and teacher at Departamento Acadêmico de Disciplinas Gerais - CEFET-MG. Main areas of interest: reading; cognition; computers and education; hypertext.

Sidney Carlos de Melo, Industrial Engineering undergraduate - CEFET-MG, under a CNPq grant.