

**The English Component of the 2002
Entrance Examination of the Federal University
of Amazonas:
An Analysis of its Reliability and Validity**

El componente de inglés del examen de admisión de la Universidad Federal del
Amazonas: Un análisis de su confiabilidad y validez

Nilton Hitotuzi*

Federal University of Amazonas
nhitotuzi@uea.edu.br

The study reported in this paper was: to estimate the reliability and validity of the English test administered to the candidates sitting the 2002 entrance examination for undergraduate courses at the Federal University of Amazonas (UFAM/2002). The test's reliability was estimated in the form of the correlation coefficient found with the test-retest method, and its validity was established on the basis of face validity and content validity. The sample population consisted of students completing their education at a high school in 2002. The results indicated that, albeit significantly generalisable to the entire population of examinees, the UFAM/2002 English test had neither reliability nor face or content validity.

Key words: English-Examination, Research-Reliability, Research-Validity

El estudio reportado en este documento tuvo como finalidad estimar la confiabilidad y validez de un examen de inglés administrado a los candidatos del año 2002 al examen de admisión para cursos de pregrado en la Universidad Federal de Amazonas (UFAM/2002). La confiabilidad del estudio se estimó mediante la correlación de coeficiente encontrado con el método test-retest y la validez se estableció a partir de la validez de forma y de contenido. La muestra poblacional consistió en un grupo de estudiantes que estaba culminando sus estudios de secundaria en el 2002. Los resultados indicaron que a pesar de la generalización significativa a toda la población de examinados, el examen de inglés de la UFAM/2002 no tuvo ni confiabilidad ni validez en su contenido.

Palabras claves: Evaluación-Inglés, Confiabilidad-Investigación, Validez-Investigación

* **Nilton Hitotuzi** holds an MA in TEFL/TESL from the University of Birmingham and a BA in Liberal Arts from the Federal University of Amazonas. Currently he holds a lecturing position at the State University of Amazonas. He is interested in Teaching Methodology and Pronunciation Teaching.

I. INTRODUCTION

“There will always be some tension between reliability and validity. The tester has to balance gains in one against losses in the other” (Hughes, 1989, p. 42).

The first institution of higher education established in Brazil, the Federal University of Amazonas (UFAM), was founded in 1909 in the State of Amazonas (*UFAM-PSM2001 - Manual do Candidato*, 2000, p. 1, 7). UFAM is home to twenty thousand students attending undergraduate courses (US) in the main city, Manaus, and in other municipalities of the State. Recently the University has increased its intake from 1,785 to 1,941 students (*UFAM-PSM2003 – Manual do Candidato*, 2002, p. 1, 6). Nonetheless, EFL teachers have informally argued that the English component of the entrance examination to UFAM technically has no influence on the selection of candidates for its undergraduate courses: candidates can completely fail the test, and still enter the University (*UFAM-PSM2002 – Manual do Candidato*, 2001, p. 4). Seeing that the rationale behind this policy may entail an array of sociocultural, political and economic factors, any attempt at verifying such expostulation would be likely to adopt a bottom-up approach. Thus it was decided that the reliability and validity of the test merited priority in the investigation, being the English test administered to the candidates sitting the 2002 entrance examination for undergraduate courses of this University (UFAM/2002) a good starting point. However, considering the involved dimensions of these concepts

(Bachman, 1990), the results herein described are tentative at best.

The reliability of the UFAM/2002 English test was estimated in the form of the correlation coefficient found with the test-retest method, and its validity was established on the basis of face validity and content validity. The sample population consisted of students completing their education at a secondary school in 2002. The results indicated that, albeit significantly generalisable to the entire population of examinees, this test had neither reliability nor face or content validity.

2. LITERATURE REVIEW

Traditionally, tests have been considered the commonest measurement instruments used for assessing learners' language abilities (Bachman, 1990; Cohen, 1994; Harris, 1969; Henning, 1987; Hughes, 1989). Carroll (1968: 46, quoted in Bachman, 1990, p.20) defines a test as “a procedure designed to elicit certain behaviour from which one can make inferences about certain characteristics of an individual”.

However, when devising tests for language learners, test developers should consider a number of crucial aspects and conditions such as the focus of the test in terms of the language abilities being tested, the method employed, and the issues of reliability, validity and feasibility. In this brief review, each one of these facets will be expanded on with a focus on learners of English as a foreign or second language.

2.1. Testing Learners' Language Abilities

An important factor in determining the language abilities to be tested is the purpose of the assessment. Cohen (1994: 23) suggests three general purposes, and twelve specific purposes (Figure 1), but maintains that "the major split is often between *proficiency* tests intended for administrative purposes and *achievement* tests for

assessment of instructional results". The marking difference between these two types of tests is that the former is 'theory-based' (e.g., TOEFL, Cambridge CPE, and ELTS), and the latter is 'syllabus-based' (Bachman, 1990; Henning, 1987). Moreover, while proficiency tests may also be administered to assess students' achievement in a particular course, achievement tests can rarely be suitable for assessing general learner proficiency (Cohen, 1994).

General purpose of the assessment	Specific reason for the assessment
Administrative	<ul style="list-style-type: none"> general assessment placement certification exemption promotion
Instructional	<ul style="list-style-type: none"> diagnosis evidence of progress feedback to the respondent evaluation of teaching or curriculum
Research	<ul style="list-style-type: none"> evaluation experimentation knowledge about language learning and language use

Figure 1. The Purpose of Assessment (Cohen, 1994, p. 23)

Another vital distinction between achievement tests and proficiency tests is that, while there seems to be no irreparable damage to examinees when results diverge greatly from their actual language competence (either due to temporary

instability on their part or inadequacy in achievement test construction), the profile and number of examinees and examiners, examiners' limitations regarding candidates, and the purpose of the assessment make a large margin of error by both examinees and

test developers in respect to proficiency tests unavoidable (Harris, 1969).

In the last part of the 1960s, focusing mostly on proficiency tests and achievement tests, Dr. David P. Harris proposed a six-way division of language test contents, namely: listening, speaking, reading, writing, grammatical structure, and vocabulary. The first four categories being labelled *complex skills*, and the last two 'components which may merit separate testing by virtue of their general importance and applicability to all four skills areas' (Harris, 1969).

While no substantial change has affected the range of categories in language tests since Harris's six-way division, there have been major developments in the mapping of communicative language abilities as a result of a number of research studies conducted in this area. This is amply exemplified by the comparison between Harris's (1969) and Cohen's (1994)

descriptions of components of oral communicative ability. For one thing, the former suggests that either four or five components are generally recognized in analyses of the speech process:

1. *Pronunciation* (including the segmental features –vowels and consonants– and the stress and intonation patterns)
2. *Grammar*
3. *Vocabulary*
4. *Fluency* (the ease and speed of the flow of speech). (Harris, 1969, p. 81).

Whereas the latter divides components of communicative ability into four different categories: (sociocultural ability, sociolinguistic ability, grammatical ability, and strategic ability (Figure 2), in passing, it seems worth observing that these categories characterise some of the essential elements of communicative competence (Stern, 1992; Bachman, 1990)).

Sociocultural ability	The appropriateness of the strategies selected for realising communicative functions in given contexts, taking into account (1) the culture involved, (2) the age and sex of the speakers, (3) their social class and occupations, and (4) their roles as status in the interaction.
Sociolinguistic ability	The linguistic forms that respondents use to express the intent of the communicative act (e.g. the grievance in a complaint).
Grammatical ability	Vocabulary, morphology, syntax, and phonology.
Strategic ability	Following Bachman's (1990: 100) suggestion, Cohen divides this ability into three components: (1) set of communicative goals, (2) retrieval of relevant items from language ability, and (3) planning of use of items retrieved.

Figure 2. Cohen's Communicative Ability (Cohen, 1994, p.10- 11)

For another thing, while Harris's view of oral proficiency seems to focus strictly on surface structures of the target language, Cohen clearly recognises other important aspects involved in communication, such as exophoric relationships (Brown and Yule, 1983), schemata (Bransford et al., 1984; Steffensen and Joag-Dev, 1984), and strategy (Bachman, 1990).

Bachman (1990) also dismisses the 'skill/component model' (to put it in his words) as being seriously limited. His (1990) detailed mapping of language abilities susceptible to testing underscores the increase in research achievements in the field towards the end of the millennium (Figure 3). This seems to suggest that assessing learners' language abilities, be it for administrative or instructional purposes, requires the development of tests so comprehensive as to embrace as much of the vast range of language competencies catalogued to date as possible.

2.2. Testing Methods

There is a large variety of language testing methods (LTM) especially designed for assessing each and every language ability (For a detailed description of LTM, see Harris,

1969; Cohen, 1994; Henning, 1987; and Bachman, 1990). Bachman (1990), for one, seems to have a broad view of LTM. So much that, in his 'framework of test method stet' (Bachman, 1990, p. 116), aspects, other than the type of test administered, are considered important variables influencing test results. One such is *time allocation*: examinees not used to working under pressure may panic and thus score very little on their exams if they are told that they have, say, 42 seconds to answer a question, as is typical of *speed tests* (Harris, 1969; Pearson and Buffa, 1994). The nature of the language, in terms of *vocabulary*, for example, is another influential aspect in test results: tasks having specialized jargon, for instance, may be too daunting for test takers. Finally, the format of the test both in terms of input and output should be yet another concern for test developers. This is the case when examinees are asked to respond in writing to listening comprehension tests, for example. If, because of physical debility, psychological problems, or other latent factors, they have a short retention span, they will fail the test, not so much for failing to understand the target-language items, but because of memory problems.

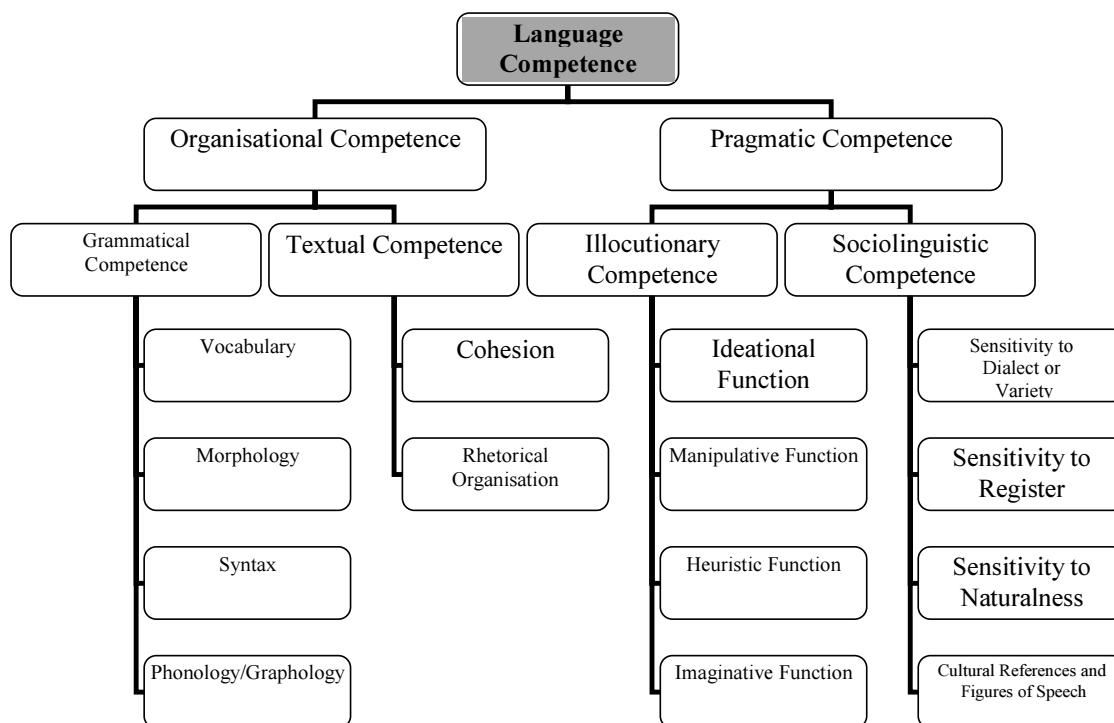


Figure 3. Bachman's mapping of language abilities susceptible of testing (Bachman, 1990, p. 87)

2.3. The Matter of Credibility

Notwithstanding the existence of a number of types of tests and a variety of aspects that can distinguish one test from another, there are some specific characteristics inherent in a test that make it reliable, valid, and feasible.

2.3.1. Reliability

Reliability, in simple terms, is the precision of a measuring instrument (Bachman, 1990; Cohen 1994; Harris, 1969; Henning, 1987; Hughes, 1989). Other things being equal, as much as there are various types of such instruments, so are the ways in which their reliability can be estimated. One way the reliability of a test,

for instance, can be established is by finding its *reliability coefficient* (RC). Harris (1969) “argues that while “a reliability quotient of 1.00 would indicate that a test is ‘perfectly’ reliable, a quotient of zero would denote a complete absence of reliability.” Testing experts make the point that the *test-retest* method (TRM) is a practical way of finding RCs (Henning, 1987). It consists of the administration of the same test to the same subjects twice, and its RC can be obtained through the Pearson Product Moment coefficient of correlation (PPM) formula (Henning, 1987):

$$r_{tt} = r_{x,y} = \frac{N\sum XY - \sum X\sum Y}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$$

where, r_{tt} = the reliability coefficient using the TRM

$r_{x,y}$ = the correlation of the scores of the number of respondents (N) of the first administration (X) with those of the second administration (Y) of the same test.

In statistics books, one can find detailed explanations on how to use this formula on paper. Fortunately, with the help of personal computers or certain types of calculators, precious time can be saved, and laborious work can be avoided when determining r_{tt} (Owen et al., 1997).

Practical as the TRM may be, one should be quick to recognise its inadequacy to account for the various facets posing a threat to test reliability. Bachman (1990), for instance, maintains that the computation of reliability by this method can be undermined by two factors: *practice effect* and *change in ability*. The former occurs if, when taking the test for the second time, examinees become acquainted with the questions or the test format. The latter is characterised by an increase or decrease in respondents' proficiency in the target language. This usually results from broadening the time span between the two administrations as an attempt to avoid practice effect. On account of that, Henning (1987) suggests 'an interval of no more than two weeks' following the administration of the first test.

Other variables that can jeopardise test reliability are examinees' psychological state and physical health (Cohen, 1994). Normally one would not be expected to perform well on a test on the day of their mother's funeral, or when having a splitting

headache, for example. Similarly, administrative environment may also lead to inconsistency of test results (Henning, 1987). Overcrowded rooms, lack of proper ventilation, pneumatic hammers rattling outside test centres, room acoustics (in the case of verbal instructions or listening comprehension tests) are some examples of administrative elements that can pose a threat to test reliability. Furthermore, *raters'* estimates of examinees' language ability may be subjective (Bachman, 1990), which is another important variable affecting scores. Many would agree that this may be even more evident with components such as written compositions and interviews. Perhaps, the frequent use of multiple-choice test formats characterises an attempt at minimising subjectivity (Bachman, 1990). Finally – no suggestion of being exhaustive intended – the inadequacy of the sampling of tasks is bound to render a given test unreliable. As a countermeasure to this problem, Harris (1969) maintains that, generally, the more samples of examinees' performance are taken, the more reliable will be the assessment of their language competency.

2.3.2. Validity

Testing specialists elaborate on a considerable number of types of validity (Bachman, 1990; Cohen, 1994; Harris, 1969; Henning, 1987; Hughes, 1989), from which only four types will be examined in this paper, namely, *face validity*, *content validity*, *construct validity*, and *criterion-related validity*.

The first type, face validity, is established on the basis of impression; therefore,

without any empirical evidence, one assumes a given test measures what it is supposed to. However, although 'this is not validity in the technical sense,' as Harris (1969) suggests, one significant aspect of face validity relates to how "testees" may translate teacher or school policy on the basis of test formats. Thus, tests that *look* silly, for instance, are likely to undermine teachers or institutions' credibility.

The second type, content validity, relates to the representativeness of the knowledge a given test is designed to measure. If, for example, a group of L2 students have covered adjective building involving prefixes and suffixes in a given course, and the test they sit is restricted to adjective building only in terms of suffixation, one can hardly argue the content validity of such a test. Likewise, if the test includes questions involving items such as *a two-hour journey* and *caramel-cum-butterscotch-cum-coffee-cum-toffee spread* (Kay and Jones, 2001, p. 76), which are types of adjectives that have not yet been taught in the imaginary course, the test is said to have no content validity. Nevertheless, a satisfactory paradigm for judging whether or not a test has content validity seems to be a *specification* of the language abilities which are bound to be covered in a course. Hughes (1989: 22) maintains that 'such a specification should be made at a very early stage in test construction.' He also argues that although not all of the language items listed in the specification may appear in the test, this typology 'will provide the test constructor with the basis for making a principled selection of elements for inclusion in the test'.

Briefly, the third type, construct validity, is verified by testing given abilities or sub-abilities in various ways. If, as a result, test scores correlate highly in line with the researcher's theory of what the frameworks of those abilities or sub-abilities are, then the test is said to have construct validity (Cohen, 1994; Henning, 1987; Hughes, 1989). Nevertheless, while arguing that there is no sufficient evidence for determining construct validation in test-score correlations, Bachman (1990) draws on Messick (1988, cited in Bachman, 1990, p. 258) to suggest other 'types of empirical evidence': "These may include any or all of the following: (1) the examination of patterns of correlations among item scores and test scores, and between characteristics of items and tests and scores on items and tests; (2) analyses and modelling of the processes underlying test performance; (3) studies of group differences; (4) studies of changes over time, or (5) investigation of the effects of experimental treatment" (Bachman, 1990, p. 258).

Finally, the fourth type of validity, criterion-related validity, embraces two other types: concurrent validity and predictive validity. The determination of the first is on the basis of the relationship between a given test and a construct-validated external criterion measure administered concurrently. The second is normally conveyed as a correlation coefficient with some future criterion measure, as Cohen exemplifies: "As an example of predictive validation, a language aptitude test may be validated by a test of a student's achievement in the language class in which the student was

placed on the basis of the aptitude test” (1994: 39- 40).

There are some caveats, though, involving these varieties of criterion-related validity. In the case of concurrent validity, even if the criterion measure has construct validity and highly correlates with a given test, no guarantee is given whether or not test scores, indicators of a given ability, are not influenced by other abilities (Bachman, 1990). On the other hand, tests developed to predict future performance cannot be considered ‘valid indicators of ability’ (Bachman, 1990). In a sentence, the author underscores the limitations of criterion-related validity: “information about criterion relatedness – concurrent or predictive – is by itself insufficient evidence for validation” (Bachman, 1990, p. 253).

Reliability versus Validity

Generally speaking, reliability and validity are closely related concepts and useful tools for test analysis (Bachman, 1990, Hughes, 1989). However, they can on certain situations be completely unconnected. Thus, for example, a test that is considered reliable because of high correlation of scores may be dismissed as invalid due to lack of consistency in terms of its purported measurement objective. In other words, all valid tests are reliable, but not all reliable tests are valid (Henning, 1987).

2.3.3. Feasibility

While reliability and validity are regarded by testing experts as essential measurement

qualities, feasibility seems to be, as Harris (1969) puts it, another *sine qua non*, without which the other two are rendered useless. This is mainly due to such things as administrative costs, availability of equipment, manpower, and appropriate facilities. While one cannot overlook other potential factors that may render a test unviable, in the last analysis, it seems the less costly the more feasible a test becomes, notwithstanding the danger of sacrificing quality for the sake of economy.

3. METHOD

3.1. Subjects

The subjects were two groups of secondary school (SS) students (N 45) in the 17 to 20 age bracket. When the investigation was conducted, these 45 students were sitting their SS final examination at *Colégio Amazonense Dom Pedro II*, a State school for primary (EF) and secondary (EM) education¹ in the State of Amazonas (Brazil).

3.2. Materials

For data collection, ninety copies were made of the UFAM/2002 English test (Appendix A). Additionally, a copy of the *candidate handbook* (CH) was examined in terms of whether or not that particular English test sampled adequately the abilities (or sub-abilities) it purported to measure (Figure 4).

¹ Ensino Fundamental (EF) and Ensino Médio (EM) respectively .

<p>Study guide for the foreign language test (Spanish, French or English)</p>	<p>For the foreign language test (FLT), there is no specific study guide. The objective of this test is just to verify whether or not the candidate has basic knowledge of the selected foreign language.</p> <p>Therefore, the FLT will involve the experience the candidate may have acquired in secondary school (<i>ensino médio</i>). It will include reading comprehension, basic vocabulary, and grammatical structure of the target language.</p>
<p>Guia de estudo para a prova de língua estrangeira (Espanhol, Francês e Inglês)</p>	<p>Para a prova de Língua Estrangeira, não há um programa específico. O que se pretende é apenas verificar se o candidato possui os conhecimentos fundamentais do idioma que escolheu.</p> <p>Assim sendo, a prova procurará levar em consideração a experiência que o Ensino Médio, geralmente, proporciona aos estudantes. Ela versará sobre o domínio de um vocabulário básico e sobre a estrutura gramatical da língua.</p>

Figure 4. PSM/2002 Candidate Handbook: study guide (UFAM-PSM2002 – Manual do Candidato, 2001, p. 9) (My translation of the 'study guide' in the PSM/2002 Candidate Handbook)

3.3. Procedure

On the basis of the test-retest method, the UFAM/2002 English test was administered twice to the forty-five students concluding SS at *Colégio Amazonense Dom Pedro II*. In agreement with Henning's (1987) suggestion on time lag between test administrations, the first testing was held on 18 November 2002, and the retesting was sat on 26 November 2002, seven days after the first administration. The test was then rated on a scale of zero to ten marks, and its reliability was estimated on the basis of the PPM formula (see section 2.3.1).

As more than twenty thousand candidates sat the 2002 entrance

examination to UFAM, 45 respondents is obviously an insignificant sample of the entire population of examinees. Therefore, so that 'the *level of significance*, or the *p-value* of the statistic' (Henning, 1987) could be verified, the *Critical Values of the Pearson Product-Moment Correlation Coefficient* (reproduced in Henning, 1987, p. 170) was used (Table 1, in Appendix B). Additionally, on the grounds that the correlation between the scores of respondents representative of the entire population aforementioned should be positive, a *one-tailed* test of significance was conducted. Yet, if a positive or negative correlation would be equally acceptable, then a *two-tailed* test of significance would be adequate (Henning, 1987).

The content validity of the test was verified by comparing the UFAM/2002 English test both with the specifications for the test outlined in the CH (Figure 4) and the parameters for the teaching of modern foreign languages (PCN) found in the

National Curriculum Parameters for Secondary School –Modern Foreign Languages (*Parâmetros Curriculares Nacionais: ensino médio – PCN, 1999, p. 147- 153; see Figure 5).*

<p>National Curriculum Parameters for Secondary School: Modern Foreign Language competencies and abilities to be developed</p>	<p>Representation and communication</p>	<ul style="list-style-type: none"> • Choose the adequate register and vocabulary for the occasion when communication occurs. • Use coherence and cohesion devices in the oral/written production. • Use verbal and non-verbal strategies to compensate for lack of ability in the target language both in production and reading. • Know and use modern foreign languages as a means to have access to other cultures, and other social groups.
	<p>Investigation and comprehension</p>	<ul style="list-style-type: none"> • Understand the various ways a given expression can be interpreted on the basis of socio-cultural aspects. • Analyse verbal language expression resources considering the nature, function, organisation and the structure of the target language (TL). Such analysis should be carried out on the basis of the proper conditions for production/reception of the TL (intention, time, place, interlocutors, and available technological resources).
	<p>Socio-cultural Contextualisation</p>	<ul style="list-style-type: none"> • Distinguish linguistic variants. • Understand to what extent oral/written texts reveal speaker traits such as behaviour, thought, and mood.

Competências e habilidades a serem desenvolvidas em Língua Estrangeira Moderna	Representação e comunicação	<ul style="list-style-type: none"> • Escolher o registro adequado à situação na qual se processa a comunicação e o vocábulo que melhor reflita a idéia que pretende comunicar. • Utilizar os mecanismos de coerência e coesão na produção oral e/ou escrita. • Utilizar as estratégias verbais e não-verbais para compensar as falhas, favorecer a efetiva comunicação e alcançar o efeito pretendido em situações de produção e leitura. • Conhecer e usar as línguas estrangeiras modernas como instrumento de acesso a informações a outras culturas e grupos sociais.
	Investigação e compreensão	<ul style="list-style-type: none"> • Compreender de que forma determinada expressão pode ser interpretada em razão de aspectos sociais e/ou culturais. • Analisar os recursos expressivos da linguagem verbal, relacionando textos/ contextos mediante a natureza, função, organização, estrutura, de acordo com as condições de produção/recepção (intenção, época, local, interlocutores, participantes da criação e propagação de idéias e escolhas, tecnologias disponíveis).
	Contextualização sócio-cultural	<ul style="list-style-type: none"> • Saber distinguir as variantes lingüísticas. • Compreender em que medida os enunciados refletem a forma de ser, pensar, agir e sentir de quem os produz.

Figure 5. The National Curriculum Parameters for Secondary School – Modern Foreign Language (Parâmetros Curriculares Nacionais: ensino médio, 1999, p. 153. My translation).

4. RESULTS OF CORRELATIONAL AND DOCUMENTAL ANALYSES

4.1. Estimating the Reliability of the UFAM/2002 English Test

On the basis of the PPM, the reliability coefficient of the UFAM/2002 administered to the forty-five respondents at *Colégio Amazonense Dom Pedro II* was $r_{tt} = 0.57$ (Table 2, in Appendix B). Furthermore, assuming that the sample of respondents was potentially comparable with samples chosen at random from the real universe of UFAM/2002 candidates, and that the correlation of their scores should be positive, the one-tailed level of significance was established. Following Table 1 (Appendix B), the value of the PPM found with the sample would have to exceed 0.2428 at a desirable level of significance for one-tailed tests ($p < 0.05$) so that generalisation of the UFAM/2002-candidate population could be made. Since r_{tt} emerged as 0.57, this correlation coefficient is considered significant and thus it could be generalised to that population if the respondents were a random sample of it. Additionally, to achieve a level of significance of $p < 0.01$, which represents 99 percent confidence in the generalisation (Henning, 1987), the PPM of the UFAM/2002 sample would have to exceed 0.3384. As the correlation coefficient found with the sample did exceed those figures, it can be considered highly significant, thus highly generalisable to the entire population, were the samples chosen from the real universe of UFAM/2002 candidates.

4.2. Estimating the Validity of the UFAM/2002 English Test

For the sake of operational viability, the validity of the UFAM/2002 was checked against the two simpler types of validity: face validity and content validity. The analysis based on 'impression' has revealed important facts about this all-reading comprehension test.

Following the typical multiple-choice format of four alternatives (Cohen, 1994), the UFAM/2002 English test was based on two texts: the first, a 211-word excerpt, apparently part of a newspaper feature article; the second, a 290-word passage, which might have been quoted from a magazine. Furthermore, the former text was followed by three items and the latter by two items. However, as one of the items of the first text had eventually to be declared void (there was no correct alternative in it), the test was reduced to four items, most of which were badly written.

For one thing, besides the fact that the respondents' native language was the language of response (Bachman, 1990), some distractors were either clearly absurd, or conflicting with one another:

(26) According to the text:

- A. All of them are the same size, and larger than Earth.
- B. There is no certainty about the discovery of these planets.
- C. There is certainty only about the existence of three new planets.
- D. There's certainty about the discovery of eight new planets.

- E. All of them can be seen from Earth with the naked eye.

(UFAM – PSM2002, Prova I, 2001, p. 5; see Appendix A for complete translation of items – my translation).

B and E represent the only real choices in this item for two reasons. One is that choice A can be instantly dismissed without reference to the excerpt (Appendix A), since the idea of a planet being ‘the same size’ as another seems as untenable as the existence of the perfect circle. The other is that choices B, C, and D eliminate one another. Thus, one might choose B on the optimistic assumption that no celestial body that can be seen ‘with the naked eye’ from Earth has escaped astronomers’ records.

For another thing, other distractors could also be easily eliminated particularly owing to bad semantics:

(28) The text mentions *brown dwarfs*. Brown dwarfs are:

- A. Small stars similar to Earth, but which have their own interior nucleus (*sic*).
- B. Small planets which have little interior nuclear power source.
- C. Stars which have nuclear mass interior (*sic*) above normal.
- D. Celestial bodies without a nucleus of their own (*sic*), but which have large mass.
- E. Celestial bodies lacking interior nuclear power source, and having less mass than planets.

(UFAM – PSM2002, Prova I, 2001, p. 5; see Appendix A for complete translation of items – my translation)

The prime example of this is the use of the word *próprio* (own) in choices A and D, which admits the assumption that a celestial body can ‘borrow’ other celestial bodies’ nucleus. Similarly, the word *interior* (interior) combined with *núcleo* (nucleus) in choice A generates blatant redundancy in Portuguese. Choice C is exemplary of the nonsensical quality of the item. Additionally, while choice E is the correct answer, its opposition to choice B seems to undermine test credibility (Harris, 1969). These facts appear to be crucial to determining the face validity of the UFAM/2002.

The comparison between the content of the UFAM/2002 with the study guide outlined in the candidate handbook and the parameters in the PCN manual has also provided valuable data for estimating the content validity of this test.

In the CH study guide it is claimed that “For the foreign language test (FLT), there is no specific study guide. The objective of [the UFAM/2002 English test] is just to verify whether or not the candidate has basic knowledge of the selected foreign language. Therefore, the FLT will involve the experience the candidate may have acquired in secondary school (*ensino médio*). It will include reading comprehension, basic vocabulary, and grammatical structure of the target language” (UFAM-PSM2002 – *Manual do Candidato*, 2001, p. 9 –My translation; Figure 4).

Nevertheless, irrespective of the sentence structuring system and the load of Latinate words typically found in academic texts written in English, which in passing favours Portuguese speakers a great deal, no meticulous examination is required to notice that the two texts selected for the UFAM/2002 are patently loaded with somewhat involved syntactic and morphological structures, and complex lexical items, thus demanding from the candidates far more than 'basic knowledge' of the target language. In terms of lexis, for instance, words and expressions such as 'stars', 'hunt for', 'scan the heavens', 'pull' 'brown dwarfs', 'lack', in the first text; and 'hard', 'ought to', 'set up', 'entire', 'too early' in the second seem to be challenging for most Brazilians whose English was learnt solely at regular schools.

Regarding the complexity of syntactic and morphological structures found in the UFAM/2002 English test, an example from the first text is:

European astronomers on Monday reported detecting signs that eight planets, some of them larger than Jupiter, may be orbiting stars outside our own solar system. (UFAM – PSM2002, Prova I, 2001, p. 4; Appendix A)

The inexistence in the Portuguese language of a structure similar to the type underlined here which would fit this particular co-text seems to contribute to potential misinterpretation of the sentence by Brazilian learners of English. In this environment, the English *-ing* form would most certainly be replaced by the Portuguese 'compound personal infinitive' (*infinitivo*

pessoal composto), e.g., *terem detectado sinais* (Cunha, 1986, p. 387).

In the third paragraph of the second text, the sentence "Because everyone has Internet access, older people participate as much as younger ones, and everyone can visit distant libraries and museums as easily as nearby ones" (UFAM – PSM2002, Prova I, 2001, p. 5; Appendix A) appears to be a good example of complex sentence structuring, mainly concerning the use of comparatives. A more complicating factor, however, seems to be the instances of test-writing negligence, which is transparent in this same paragraph: 'where' and 'and' are written as 'whwre' and 'end' respectively. This inattentiveness is indeed more markedly noticed in the construction of item two following the first text, which was declared void for lacking a correct alternative.

About the 'experience' the candidate may have acquired in EM, this is a rather involved issue, inasmuch as, regardless of the PCN (Figure 5), the EFL classroom reality, especially in state schools, seems to indicate that students finishing EM have only superficial knowledge of English – this is yet to be empirically investigated though.

5. DISCUSSION

5.1. The Reliability of the UFAM/2002 English Test

The reliability coefficient of the UFAM/2002 found with the 45 respondents ($r_{tt} = 0.57$) through the TRM indicates an extremely low level of reliability. Harris (1969) maintains that tests designed to

select candidates 'should have a reliability quotient of at least 0.90, and preferably even somewhat higher'. Nevertheless, as he points out, the limitations of this method should be considered when making decisions on the basis of correlation coefficients. In the case of these respondents, for instance, I noticed that some of them returned the test far sooner than the estimated time for them to answer it. Considering the degree of difficulty of the passages, it was established that the respondents should be able to answer the test in twenty minutes. However, some of the respondents returned the test in about seven minutes after they were given it. This phenomenon seems to constitute a different source of inconsistency other than the classic ones: 'differential practice effect and differential changes in ability' (Bachman, 1990, p. 182). Perhaps these respondents' attitude is an indication of disappointment. It is possible that they may have felt that the English they had been learning so far at school was not enough for them to understand the texts in the UFAM/2002 English test. Hughes (1989) seems to endorse this view: "Now it might be argued that to base test content on objectives rather than on course content is unfair to students. If the course content does not fit well with objectives, they will be expected to do things for which they have not been prepared" (1989: 12).

5.2. The Validity of the UFAM/2002 English Test

In the light of the evidence provided in the previous section, it seems reasonable to

say that the UFAM/2002 English test lacks both face validity and content validity. In regard to the former, although the word-count of the two selected texts fluctuates within the range of acceptable text length, both the number of texts, and items per text are far below testing experts' prescription for this type of test. Harris (1969) suggests about five passages, each ranging from 100 to 250 words, and followed by four to seven items or more. Furthermore, the badly written items and the two spelling problems detected in the second passage also undermine the face validity of the UFAM/2002 English test. At this point, it is important noting that the *lead* in (26) is rather vague, thus requiring that 'them' be replaced by a noun phrase in choice A (or A could be swapped for B), so that A could be clearly understood in isolation. This appears to be endorsed by Harris (1969: 62): "The vocabulary and syntax of the items should be kept as simple as possible so that the real problem is the interpretation of the *passage*, not of the questions that are asked about it".

Importantly, on the basis of Harris's (1969) 'advice on item writing', a suggestion for items following the first passage of the UFAM/2002 English test is outlined in Appendix C.

To consider the next point, the content validity of the UFAM/2002 English test, one has to partly assume a speculative stance, inasmuch as a study is yet to be conducted on whether or not the PCN are in fact followed in the EFL classroom at both private and state schools in the State of Amazonas. Nevertheless, when collating the CH study

guide with the two texts of the UFAM/2002 English test there can hardly be any doubt about the gap between what is stated in the CH, and the actual level of proficiency required by those texts. This nonetheless seems to be mitigated by both the choice of the candidates' native language as the language of response and the unchallenging quality of the items.

Albeit the suggestion in Appendix C seems to offer more reliable items for the first passage of the UFAM/2002 English test, many would be likely to agree that the test would still have no validation considering the argument on the comparison between the UFAM/2002, the CH study guide and the PCN. As testing experts suggest, a test is valid only insofar as it constitutes a sample of what it purports to measure (Bachman, 1990; Cohen, 1994; Harris, 1969; Henning, 1987; Hughes, 1989).

6. CONCLUSION

According to the results presented in this paper, the expostulation of ELT teachers working at state schools in Manaus was confirmed: the English component of the entrance examination to the Federal University of Amazonas administered in 2002 was neither reliable nor valid. In other words, the UFAM/2002 English test was considered dispossessed of basic requirement to be qualified as a serious screening test of English. However, the weight of this conclusion is put into question on account of sample reliability. Thus, to validate the findings reported here, the repetition of this investigation under more reliable conditions appears to be necessary.

Nevertheless, as tentative as this study may be, it is felt that, once confirmed, these results seem to have serious implications particularly for the English undergraduate course (EUC) of this University whose unstated required level of proficiency in the target language is at least two years of comprehensive and continuous study – EUC freshmen have to deal with *New Interchange 2* (Richards and Sandy, 1998) in a classroom where English is supposed to be the language of instruction. Unfortunately, considering the length of the EUC, the adoption of such a tacit policy is understandable: it is rather unlikely that beginners could become advanced learners in only eight terms of study. One can suggest two alternative solutions to this dilemma though.

One solution would be the addition of two more terms to the EUC, so that freshmen could start studying English as 'real' beginners. Hopefully, they would achieve advanced level proficiency by the end of the course. The other solution would be the development of a reliable and valid screening test for candidates applying for the EUC. However, this measure could be dismissed by both the University and the local community for different reasons. For one thing, lack of manpower and money to operate the logistics involved in the administration of such a test may potentially be UFAM's strong claims. For another thing, this new policy may attract widespread criticism from the Amazonian community. For instance, one might argue against the discriminatory quality of the process, claiming that the seven years of EFL learning at state schools (as well as private schools) fail to prepare candidates to succeed in this

type of test. From this perspective, then, it is likely that only those who could afford to study in a language institute would be apt to attend the EUC at UFAM.

Since no empirical study has been conducted on this issue to date, the suggestion that Amazonian students cannot achieve an intermediate level of proficiency in English after having studied this subject for fourteen terms at state and private schools can only be speculative. However, the results presented in this paper seem to lead to a number of questions about the role of English (or any other modern foreign language) in the Brazilian formal educational system, which may merit further research.

For one thing, one might want to investigate, for instance, the extension to which the PCN are based on sound principles of foreign language teaching, or the validity of the assumptions underlined in Figure 5 about the level of proficiency of EM students. Another concern is how closely the PCN are followed in the EFL classroom of state and private schools. One way of verifying this would be through an analysis of the ELT material used in these schools. This might determine whether or not the EFL classroom of EF is conducive to the development of the target language in EM.

For another thing, assuming the English component of the entrance examination to the Federal University of Amazonas were decisive to admission not only to the EUC, but also to the other undergraduate courses offered by this University, how would English compare with other subjects in the EF and

EM curricula? How seriously would EF and EM students take it? What influence would these conditions have on school policy concerning the selection of EFL teachers, and adequacy of environment for EFL teaching?

REFERENCES

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bransford, J. D. et al. (1984). Learning from the perspective of the comprehender. In Anderson, J.C. and Urquhart, A. H. (eds.) *Reading in a foreign language*. London: Longman.
- Brasil. (1999). Ministério da Educação, Secretaria de Educação Média e Tecnológica. *Parâmetros Curriculares Nacionais: Ensino Médio*. Brasília: Ministério da Educação.
- Brown, G. and Yule, G. (1983). *Discourse analysis*. Cambridge: Cambridge University Press.
- Carroll, J.B. (1968). The psychology of language testing. In: Bachman, L. F. *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Cohen, A. D. (1994). *Assessing language ability in the classroom*. (2nd ed.). Boston, Mass: Newbury House.
- Cunha, C. F. (1986). *Gramática da língua portuguesa*. (11^a ed.). Rio de Janeiro: FAE.
- Harris, D. P. (1969). *Testing English as a second language*. New York: McGraw-Hill.

Henning, G. (1987). *A guide to language testing: Development, evaluation, research*. Boston, Mass: Newbury House.

Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.

Kay, S. and Jones, V. (2001). *Inside out upper intermediate: Student's Book*. Oxford: Macmillan Heinemann.

Owen, C., et al. (1997). *Testing*. Birmingham: Centre for English Language Studies (CELS).

Pearson, L. and Buffa, L. (1994). *The Princeton review: Cracking the TOEFL, 1995 edition*. New York: Villard Books.

Richards, J. C. and Sandy, C. (1998). *New Interchange 2: Student's Book*. Cambridge: Cambridge University Press.

Steffensen, M. S. and Joag-dev, C. (1984). Cultural knowledge and reading. In Anderson, J. C. and Urquhart, A. H. (eds.). *Reading in a foreign language*. London: Longman.

Stern, H.H. (1992). *Issues and options in language teaching*. Oxford: Oxford University Press.

Universidade Federal do Amazonas. (2000). *Processo Seletivo Macro – PSM2001: Manual do Candidato*. Manaus: Imprensa Universitária.

Universidade Federal do Amazonas. (2001). *Processo Seletivo Macro – PSM2002: Prova 01, Conhecimentos Gerais I*. Manaus: Imprensa Universitária.

Universidade Federal do Amazonas. (2001). *Processo Seletivo Macro – PSM2002: Manual do Candidato*. Manaus: Imprensa Universitária.

Universidade Federal do Amazonas. (2002). *Processo Seletivo Macro – PSM2003: Manual do Candidato*. Manaus: Imprensa Universitária.

APPENDIX A: THE UFAM/2002

Processo Seletivo Macro – Psm2002 Prova de Conhecimentos Gerais I

Língua Estrangeira - Inglês

(Passage one)

LEIA O TEXTO “EIGHT NEW POSSIBLE PLANETS DETECTED” E RESPONDA ADEQUADAMENTE:

WASHINGTON (Reuters) – European astronomers on Monday reported detecting signs that eight planets, some of them possibly larger than Jupiter, may be orbiting stars outside our own solar system.

These discoveries bring the number of potential extrasolar planets to more than 40. The hunt for these planets candidates has intensified in the last year, as space scientists from around the world scan the heavens for Sun-like stars and planets that might orbit them.

The newest candidates were detected by astronomers working with the European Southern Observatory’s La Silla observatory in Chile. The scientists are based in Geneva, Switzerland.

None of the planet candidates have ever been seen by humans, but scientists believe they are there because of the gravitational pull they exert on the stars they orbit.

Two of the eight new candidates may not be planets at all, the European scientists said in a statement, but could instead be brown dwarfs, which have a bit less mass than stars and completely lack a star’s interior nuclear power source.

Three of the new planet possibilities are about the size of Saturn or smaller, three are one to three times the size of Jupiter and two are 10 times the size of Jupiter or larger.

All of these are far larger than Earth.

26. De acordo com o texto:

- a) Todos são do mesmo tamanho e maiores do que a Terra
- b) Não há certeza quanto à descoberta desses planetas
- c) Há certeza só quanto à existência de três novos planetas
- d) Há certeza quanto à descoberta de oito novos planetas
- e) Todos podem ser vistos da Terra a olho nu.

27. Os sinais detectados indicam que: (NULA)

- a) os oito planetas seriam maiores do que Júpiter e estariam fora de nosso sistema solar.
- b) os oito planetas seriam menores do que Júpiter e estariam dentro de nosso sistema solar.
- c) os oito planetas seriam iguais a Júpiter e estariam na orla de nosso sistema solar.

- d) os oito planetas são, na verdade, maiores do que a Terra e dentro de nosso sistema solar (sic).
- e) os oito planetas são bem maiores do que a Terra e próximos de Saturno (sic).

28. O texto faz menção a "brown dwarfs". "Brown dwarfs" são:

- a) pequenas estrelas semelhantes a Terra, mas com núcleo interior próprio (sic).
- b) pequenos planetas com pouca fonte de energia nuclear.
- c) estrelas com interior de massa nuclear acima do normal (sic).
- d) corpos celestes sem núcleo próprio(sic), mas grande massa.
- e) corpos celestes desprovidos de fonte de força nuclear interna e possuem menos massa do que os planetas.

(My translation of instructions and items for the first passage)

READ THE TEXT "EIGHT NEW POSSIBLE PLANETS DETECTED" AND ANSWER ADEQUATELY:

26. According to the text:

- a) All of them are the same size, and larger than Earth.
- b) There is no certainty about the discovery of these planets.
- c) There is certainty only about the existence of three new planets.
- d) There's certainty about the discovery of eight new planets.
- e) All of them can be seen from Earth with the naked eye.

27. The detected signs indicate that: **(VOID)**.¹

- a) The eight planets would be larger than Jupiter and would be out of our solar system.
- b) The eight planets would be smaller than Jupiter and would be inside our solar system.
- c) The eight planets would be similar to Jupiter and would be bordering our solar system.
- d) The eight planets are in reality larger than Earth and inside our solar system (sic).*
- e) The eight planets are by far larger than Earth and next to Saturn (sic).*

¹ Because the existence of the eight planets is assumed in all choices, this item was cancelled.

* In Portuguese *ser* and *estar* (to be) convey different meanings.

28. The text mentions *brown dwarfs*. Brown dwarfs are:

- a) Small stars similar to Earth, but which have their own interior nucleus (sic).
- b) Small planets which have little interior nuclear power source.
- c) Stars which have nuclear mass interior (sic) above normal.

- d) Celestial bodies without a nucleus of their own (*sic*), but which have large mass.
- e) Celestial bodies lacking interior nuclear power source and having less mass than planets.

(Passage two)

LEIA O TEXTO "WILL TECHNOLOGY REPLACE SCHOOL?" E RESPONDA ADEQUADAMENTE:

Some people believe that soon schools will no longer be necessary. These people say that because of the Internet and other new technology, there is no longer any need for school buildings, formal classes, or teachers. Perhaps this will be true one day, but it is hard for me to imagine a world without schools. In fact, we need to look at how we can use new technology to make schools better-not to eliminate them.

We should invent a new kind of school that is linked to libraries, museums, science centers, laboratories, and even corporations. Corporations ought to create learning programs for schools in their area of expertise. Independent experts could give talks on video or over the internet. TV networks and local stations could develop programming about things students are actually studying in school. Laboratories could set up websites to demonstrate new technology so students could view it on the Internet.

Is this just a dream? No. Already there are several towns where (*sic*) this is beginning to happen. Blacksburg, Virginia, is one of them. Here the entire city is linked to the Internet, and learning can take place at home, at school, and in the office. Businesses provide programs for the schools and the community. The schools provide computer labs for people without their own computers at home. Because everyone has Internet access, older people participate as much as younger ones, and everyone can visit distant libraries and museums as easily as nearby ones.

How will this new kind of school change learning? It is too early to be sure, but it is very exciting to think about. Technology will change the way we learn; schools will change as well; and we will all learn something in the process.

29. Leia as colocações abaixo e assinale aquela com as quais o autor concordaria.

1. Escolas não são necessárias.
 2. As novas tecnologias irão substituir as escolas.
 3. O uso de novas tecnologias é positivo.
 4. A aprendizagem deveria se dar só nas escolas.
 5. As empresas deveriam se envolver com a educação.
- a) 1 – 3 – 4
 - b) 3 – 5
 - c) 4 – 5
 - d) 1 - 2 – 5
 - e) 3 – 4

30. Na frase “*perhaps **this** will be true one day...*” (1.º parágrafo), **this** refere-se:
- a) a um mundo sem mestres, só tecnologia.
 - b) à necessidade de acabar com as aulas e com os professores.
 - c) a não necessidade de prédios escolares, aulas formais ou professores.
 - d) à invenção de uma nova escola sem estrutura
 - e) à dominação dos prédios escolares pelos computadores

(My translation of instructions and items for the second passage)

READ THE TEXT “*WILL TECHNOLOGY REPLACE SCHOOL?*” AND ANSWER ADEQUATELY:

29. Read the statements below, and indicate the ones with which the author would agree.
- a) Schools are not necessary.
 - b) New technologies will eventually replace schools.
 - c) The use of new Technologies is positive.
 - d) Learning should take place only at school.
 - e) Companies should participate in the educational process.
- a. 1 – 3 – 4
 - b. 3 – 5
 - c. 4 – 5
 - d. 1 - 2 – 5
 - e. 3 – 4

30. In the sentence “*perhaps **this** will be true one day...*” (1st paragraph), **this** refers to:
- a) a world without teachers, only technology.
 - b) the necessity to banish lessons and teachers.
 - c) the fact that school buildings, formal lessons or teachers are not necessary.
 - d) the creation of a new school, without structure.
 - e) the supremacy of computers in school buildings.

APPENDIX B: TABLES

$df = N - 2$	Level of significance for one-tailed test				
	.05	.025	.01	.005	.0005
	Level of significance for two-tailed test				
	.10	.05	.02	.01	.001
1	.9877	.9969	.9995	.9999	1.0000
2	.9000	.9500	.9800	.9900	.9990
3	.8054	.8783	.9343	.9587	.9912
4	.7293	.8114	.8822	.9172	.9741
5	.6694	.7545	.8329	.8745	.9507
6	.6215	.7067	.7887	.8343	.9249
7	.5822	.6664	.7498	.7977	.8982
8	.5494	.6319	.7155	.7646	.8721
9	.5214	.6021	.6851	.7348	.8471
10	.4973	.5760	.6581	.7079	.8233
11	.4762	.5529	.6339	.6835	.8010
12	.4575	.5324	.6120	.6614	.7800
13	.4409	.5139	.5923	.6411	.7603
14	.4259	.4973	.5742	.6226	.7420
15	.4124	.4821	.5577	.6055	.7246
16	.4000	.4683	.5425	.5897	.7084
17	.3887	.4555	.5285	.5751	.6932
18	.3783	.4438	.5155	.5614	.6787
19	.3687	.4329	.5034	.5487	.6652
20	.3598	.4227	.4921	.5368	.6524
25	.3233	.3809	.4451	.4869	.5974
30	.2960	.3494	.4093	.4487	.5541
35	.2746	.3246	.3810	.4182	.5189
40	.2573	.3044	.3578	.3932	.4896
45	.2428	.2875	.3384	.3721	.4648
50	.2306	.2732	.3218	.3541	.4433
60	.2108	.2500	.2948	.3248	.4078
70	.1954	.2319	.2737	.3017	.3799
80	.1829	.2172	.2565	.2830	.3568
90	.1726	.2050	.2422	.2673	.3375
100	.1638	.1946	.2301	.2540	.3211

Table I. Critical Values of the Pearson Product-Moment Correlation Coefficient
(reproduced in Henning, 1987, p.170)

Table 2

APPENDIX C: NEW ITEMS FOR THE UFAM/2002

Here is a suggestion of items for the first reading passage of the UFAM/2002 English test.

After this passage, you will find five unfinished statements about it, each with five suggested ways of finishing. You must choose the one which you think fits best.

EIGHT NEW POSSIBLE PLANETS DETECTED

WASHINGTON (Reuters) – European astronomers on Monday reported detecting signs that eight planets, some of them possibly larger than Jupiter, may be orbiting stars outside our own solar system.

These discoveries bring the number of potential extrasolar planets to more than 40. The hunt for these planets candidates has intensified in the last year, as space scientists from around the world scan the heavens for Sun-like stars and planets that might orbit them.

The newest candidates were detected by astronomers working with the European Southern Observatory's La Silla observatory in Chile. The scientists are based in Geneva, Switzerland.

None of the planet candidates have ever been seen by humans, but scientists believe they are there because of the gravitational pull they exert on the stars they orbit.

Two of the eight new candidates may not be planets at all, the European scientists said in a statement, but could instead be brown dwarfs, which have a bit less mass than stars and completely lack a star's interior nuclear power source.

Three of the new planet possibilities are about the size of Saturn or smaller, three are one to three times the size of Jupiter and two are 10 times the size of Jupiter or larger.

All of these are far larger than Earth.

1. According to the information given in the passage about stars, one can argue that
 - A. they can be larger than planets such as Jupiter and Earth.
 - B. normally their gravitational field is less extensive than that of the planet Earth.
 - C. their interior nuclear power source is exhaustible.
 - D. they are especially orbited by planets.
 - E. they can be sun-like celestial bodies smaller than the Sun.

2. The writer suggests that these discoveries
 - A. evidence that celestial bodies can be precisely measured.
 - B. may be stars orbiting outside our solar system.
 - C. may be celestial bodies with narrow gravitational fields.
 - D. increase the number of extrasolar planets to more than forty.
 - E. were made from Chilean ground.

3. The scientists who discovered the celestial bodies are
 - A. Swiss.
 - B. Spanish native speakers.
 - C. North Americans.
 - D. South Americans.
 - E. Europeans.

4. According to the text brown dwarfs
 - A. are larger than Earth.
 - B. can be larger than some stars.
 - C. can be characterised by their similarity with distant planets.
 - D. can be confused with stars.
 - E. are ten times the size of Jupiter or larger

5. One may conclude from the passage that these celestial bodies
 - A. are orbiting outside our solar system.
 - B. are planets that can be confused with stars.
 - C. total eight extrasolar planets.
 - D. could be seen from the observatory thanks to modern technology.
 - E. are planets which have more mass than Earth.