SOME THOUGHTS ON
THE UNIVERSITY ENTRANCE EXAM
IN ENGLISH LANGUAGE
OF 1993 AND 1994

1) The official "Programa para 1994" envisages three sections in the exam (I: text + appreciation; II: "aplicação linguistica"; III: composition). Most university teachers would consider section II (basically, grammar) to be of equal importance to Section I (reading comprehension) or perhaps even Section III (writing skills) but in the 1ªchamada/1994 Section II only merits 11% of the marks available. There should be more balance and interaction across the sections so that the exam would approximate more closely to real language in use: "... integrative tests attempt to assess a learner's capacity to use many bits (of language) all at the same time, and possibly while exercising several presumed components of a grammatical system, and perhaps more than one of the traditionally recognised aspects of skills" [1].

2) It seems unacceptable that in the 1ª chamada/1994 only 5% of the marks were attributed to the candidates' direct/systematic knowledge of grammar, namely verb tenses (question II.2). There must be more space made available for a more comprehensive and challenging analysis of this part of the candidates' abilities. This could be achieved without resorting to testing formats or techniques with which the candidates are unfamiliar. However at the same time it should also be taken into account that "indirect tests are subject to attacks on their validity in those cases where it is possible to by-pass the ability in question and develop proficiency in the assessment task alone." [2].

---

[1] J. OLLER, quoted in WEIR, C. — *Communicative Language Testing*, Exeter, Exeter Linguistic Papers, Vol. 11, 1988, 1978, p. 4.
[2] A. KELLY, quoted in WEIR, C. — *Op. cit.*, 1978, p. 7.

3) The distribution of marks should give greater emphasis to the more objective elements of the test (it was only 15% in 1ª chamada/1993), making sure that there is a balance between the skills being tested and also allowing for an element of creativity. The exam does not deal with the skills of speaking and listening whereas, for example, the evaluation system in operation on the English language course at the Faculty of Letter in Porto attributes a 50% weighting to these skills regardless of the mode of evaluation the students opt for.

4) The exam itself should be laid out "on the page" according to the ease of manipulation/access it allows the candidate. For example, a reading text and the questions on that text should be printed "face to face". This element of test development and production should be given more importance: "Too often, institutional tests are badly typed (or handwritten), have too much text in too small a space, and are poorly reproduced. As a result students are faced with additional tasks which are not ones meant to measure their language ability. Their variable performance on the unwanted tasks will lower the reliability of the test." [3].

5) The sequence of questions should take into account more the relative importance of the marks they carry (why do the most important always seem to come last when the candidate may be rushing to finish?) in addition to being "accessible", as mentioned above.

6) As well as providing a page where the mark distribution is fully described (Why the last page, why not on the front of the test?) the marks per question should be printed after each question. Perhaps an indication of the recommended time to be spent per question could be included.

7) In the 1ª chamada/1994 of the marks awarded 49% were based on text prompted questions restricting the candidates who have other skills, for example, in creative writing or vocabulary. There should be more attention paid to this area in order to increase the *construct validity* of the test. (Construct validity is "a form of validity which is based on the degree to which the items in a test reflect the essential aspects of the theory on which the test is based.") [4]. Readers should consult Hughes A. (1989) pp. 34/37 for further reading on the notions of validity and reliability.

---

[3] HUGHES, A. — *Testing for Language Teachers*, Cambridge, Cambridge University Press, 1989, p .39.

[4] RICHARDS, J., PLATT, J.; WEBER, H. (eds) — *The Longman Dictionary of Applied Linguistics*, London, Longman, 1985, p. 61.

8) The exam should be more aimed at future potential (be more of a diagnostic test) and be less constrained by the syllabus/themes of the candidates learning experience at secondary school (be less of an achievement test). There is a strong sense in which this exam is an end of high school test (and therefore duplicating existing in-school assessment) rather than being a truly external predictive test. In short, it should look forward not backwards.

9) In the 1ª chamada/1994 question I.D (a text linked to American history) was weighted with 27% of the marks available. This question was entirely retrospective, referring back to the secondary school programme when the *real* purpose of this exam should be predictive: looking forward as to whether the candidates will succeed at university level i.e. the exam should be more clearly focused in terms of its *predictive validity.* Indeed "In a diagnostic situation it may become important not simply the degree of skill which a candidate can bring to the performance of a particular global task, but also to find out precisely which of the communicative skills and elements of language he has mastered... the deficiencies so revealed might form the input to a teaching programme..." [5].

10) While the above comments deal with problems areas of a largely general nature more specific doubts could also be raised referring to individual questions' content: in the 1° chamada/1994, question I.B.5 is almost impossible to get wrong; I.B.4 requires a personal opinion, through the writer's eyes(??); II.1 has several answers the same: "in" used for expressing time (in 1825/the summer/the eighteenth century) etc etc. It should be noted that "Knowledge of the elements of a language in fact counts for nothing unless the user is able to combine them in new and appropriate ways..." [6].

11) Greater care is required in formulating the instructions given to the candidates, both in general and in terms of specific questions (the cloze test instructions in 1ª chamada/1993 were very poorly worded as were the instructions for the sentence transformations), to put it bluntly the instructions should be "idiot-proof" if the prova is to remain reliable: "This

---

[5] MORROW, K. — *Communicative Language Testing: revolution or evolution* in BRUMFIT, C. J.; JOHNSON, K. (eds.) — *The Communicative Approach to Language Teaching,* Oxford, Oxford University Press, 1979, p. 152.
  [6] MORROW, K. — *Op. cit.,* p. 145.

is not the place to list all possible features which might make it unreliable, but examples are: unclear instructions, ambiguous questions, items that result in guessing on the part of the test takers." [7].

12) choice of question type/style is open to debate but surely there is little point in asking specific vocabulary questions when the candidates are allowed to bring dictionaries to the exam, unless the aim is to find out if the candidates can use/read a dictionary. Did all the candidates know they could bring a mono-lingual dictionary? If not the exam was not conducted under the same conditions for all the candidates. And from another more alarming point of view the information required to answer the question on The Boston Tea Party in the 1ªchamada/1994 is available in some good dictionaries on the market.

13) There is really no need to oblige the candidates to do the same task twice, for example, to write two mini-compositions. Why not just give them a choice of topics to write on once and then utilize the time/space to test a different skill in a more objective and/or communicative manner: "The tasks should give the candidates the chance to assert their communicative independence and allowance should also be made for the creative unpredictability of communication in the tasks set..." [8].

14) Much clearer marking guidelines need to be provided for the teachers marking the candidates' answers. Simply to say, give full marks to an answer which is "adequada" or "coerente e fundamentada" is not good enough. While all candidates did the same exam, not all the candidates can receive equal treatment in terms of marking unless there are clear, unambiguous guidelines.

15) The guidelines should provide more information on difficult areas such as: how to deal with varying degrees of error; how to deal with candidates who exceed word limits; what kind of cultural content or general knowledge is expected of the candidate; what is acceptable and what is not in the objective sections; what degree of spelling accuracy is expected and whether it is the same for all sections of the exam. The scoring key provided should: "specify acceptable answers and assign points for partially correct responses. ... It should be the outcome of efforts to anticipate all possible responses and have been subjected to group criticism." [9].

---

7 HUGHES, A. — *Op. cit.*, p. 3.
8 WEIR, C. — *Op. cit.*, p. 41.
9 HUGHES, A. — *Op. cit.*, p. 41.

16) The problem here has its origin in the lack of objectivity in many sections of the exam: subjective questions → subjective answers → subjective marking. The manner in which *scorer reliability* operates needs to be greatly enhanced. For example by having a thorough discussion of the marking scheme for all the scorers prior to any marking; an easier distribution of marks to follow (not jumping from 3 to 7 to 9 to 2 to 0.6). Ease of administration is crucial to reliable marking, the exam should display a higher level of efficiency: "A valid and reliable test is of little use if it does not prove to be a practical one. This involves questions of economy, ease of administration, scoring, and interpretation of results." [10].

17) Since reading comprehension carried so many marks it is important that strict guidelines are made available so as to prevent a natural bunching of marks across an equivalent range of 8 → 14 which in no way assists in creating an effective screening of candidates.

18) In 1994 there was a clear lack of correspondence between the two chamadas which created a huge disequilibrium in terms of candidate selection. This imbalance was apparent at a variety of levels: style/source of text, nature of questions, distribution of marks, cultural content (assumed or otherwise) etc etc. There was no *concurrent validity/reliabilty* in operation in the 1994 exam. There is a great deal of difference between attempting to test the general language proficiency of candidates and a test which may only "enable the level of proficiency to be assessed relative to that of other candidates who take the same test under the same conditions" [11]. Thus very little externally appliable information may be extrapolated from such an exam.

19) The fact that far fewer candidates sit the 2ª chamada allows for potential differences in terms of the actual physical conditions (seating/lighting/ventilation/etc.) in which the exam is conducted which may influence the results achieved by the candidates. In addition this situation may give rise to differences in the quality/manner of the exam supervision/invigilation that occurred on the two different dates. Conditions which facilitate cheating make any exam unreliable in the extreme. Readers should consult Hughes A. (1989) pp. 152/154 for further discussion of test administration.

---

[10] WEIR, C. — *Op. cit.*, p. 37.
[11] MORROW, K. — *Op. cit.*, pp. 148/149

20) No exam can be considered valid if the candidates are so close to each other (with no empty chair in-between) that they can copy freely. There should be a minimum set of conditions stipulated, for example the distance between candidates, so that the physical conditions in which the candidates sit the exam are as similar as possible in all rooms, at all sites.

21) Faculties can allow themselves little room to complain unless they are more willing to involve themselves more in the entire process of the formulation/production of the exam. "If a test is regarded as important, then preparations for it can come to dominate all teaching and learning activities. And if the test content and testing techniques are at variance with the objectives the course, then there is likely to be harmful backwash." [12]. The exam provides an almost unique opportunity for a Faculty to directly influence the composition of its student body and also more indirectly to influence secondary schools to teach up to the level of the exam rather than for a Faculty teach to level they are sent.

22) The exam production process needs to be started much earlier in the academic year to give it more space to have it discussed and worked on and in this manner hopefully produce an improved exam. This work should begin as soon as possible in the following academic year so that the exam corresponds more to the idea that "Tests should designed to reveal not simply the number of items which are answered correctly, but to reveal the quality of the candidate's language performance." [13].

N. R. Hurst

---

[12] HUGHES, A. — *Op. cit.*, p. 1.
[13] MORROW, K. — *Op. cit.*, pp. 145-146.