RECHARGING THE BATTERY: PLACEMENT TESTS FOR ESP STUDENTS

Lucia Pacheco de OLIVEIRA
Inés Kayon de MILLER
Márcia do Amaral Peixoto MARTINS
Maria Isabel Azevedo CUNHA
James Dean BROWN

• ABSTRACT: This paper first reviews key concepts in language testing and discusses the use of the English Language Placement Test administered at the Catholic University of Rio de Janeiro to computer science students, who are required to take one of the two ESP courses offered by the English program. The analysis of the scores on two versions of a test administered to two different groups of computer science students who took the examination in sequential semesters is presented. For the original test, item facility and item discrimination were calculated, then overall test reliability was estimated. The statistics were carefully examined and a number of the test items were eliminated, revamped or replaced in creating a revised version of the test. This modified and improved version of the test was then administered to a second group of students and the same statistics were applied once more. The results are discussed in terms of the importance of item analysis and revision for providing more consistent and accurate language testing instruments.

• KEYWORDS: Testing; teaching; EFL.

Introduction

Language tests are generally not easy to create. However, careful preparation, as well as a posterior analysis of results and revision of
questions may enhance the quality of different types of tests, making them more meaningful for teachers and students.

This paper describes the process five faculty members at the Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio) went through in improving a specific test: the English Language Placement Test administered to students from the computer science program.

Brown and Pennington (1991) have listed four categories of tests used to measure students' language maturity in various ways: proficiency tests, placement tests, diagnostic tests and achievement tests. Alderson, Krahnke & Stansfield (1987) also identify these four types of tests as the most commonly used in the field of language teaching (Brown, 1992). Briefly, these four types of tests can be described as follows:

(a) **Proficiency tests** are typically not based on any particular program's content. As such, they are designed to measure how much of a given language a person has learned or acquired without reference to a particular program and its objectives. Proficiency tests are usually used for global decisions like the admissibility of a student to a particular institution or language program.

(b) **Placement tests** are those instruments designed to match students' foreign language ability with the specific content of the courses in a program. They give insight for the placing of students into the various levels of study within a program. They do not usually generate grades and, in general, results are given as course assignments which correspond to the course(s) in which the students should enroll.

(c) **Diagnostic tests** aim at determining the learners' strengths and weaknesses with regard to the specific objectives of a course. Such tests are usually administered at the beginning of a program or in the middle of a course (as progress tests). The purpose of these tests is to help students and their teachers focus their efforts where they will be most effective.

(d) **Achievement tests** measure how students fare in relation to the content or use of certain skills from a narrower perspective. These tests, usually based on the specific objectives of a particular course, evaluate the extent to which the learners have assimilated the subject matter taught and practiced in the course.

The English program at PUC-Rio has experienced these four types of tests. The English Language Placement Test, which will be the
focus of this paper, is administered to computer science students who may be required to take at least one course in English for computer science. These English for specific purpose courses (ESP) are taught in two levels.

In order to improve the placement instrument and to make it more meaningful for students and teachers, numerous versions of this test have been designed by the English faculty at PUC-Rio. These attempts at revision have been designed to produce an instrument that can more accurately measure students' English maturity, especially in reading (with a focus on vocabulary, and reading for facts and inference).

The purpose of this paper is to describe the process of test revision and show the steps followed in improving our placement test. The faculty's goal was to systematically improve the original instrument administered in May 1992. This goal led to examination of the statistical results, especially item analyses, and the revision and re-administration of a revised version of the test in October 1992. The discussion of these processes will be organized around the following research questions:

1. How many test items in the original test need revision?
2. To what degree is the score distribution of the original test improved in the revised version of the test?
3. To what degree is reliability affected by the revision process?
4. To what degree is the final version of the test valid?

Method

Subjects

The subjects in this study were the entire population of first semester 1992 and second semester 1992 computer science freshmen at PUC-Rio. Because of their departmental regulations, they were required to take the English Language Placement Test in May and October 1992 in order to be correctly placed in English for Computer Science I or IA. A total of 44 freshmen took the original version of the placement test in May and 101 freshmen took the revised version in October 1992.
Table 1 shows the subjects' overall distribution in terms of conditions of language study, sex, immersion in an English-speaking country and experience in the Computer Science field. The two groups seem to be very similarly distributed in terms of their experience in living abroad and in their specific field of study. It was only in terms of sex and language study characteristics that the two groups showed differences.

Table 1 – Description of subjects who took the original and revised tests

<table>
<thead>
<tr>
<th>Variable</th>
<th>May 1992 (N = 44)</th>
<th>October 1992 (N = 101)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years of English Study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-5</td>
<td>56 %</td>
<td>64 %</td>
</tr>
<tr>
<td>6-10</td>
<td>38 %</td>
<td>33 %</td>
</tr>
<tr>
<td>11-15</td>
<td>4 %</td>
<td>2 %</td>
</tr>
<tr>
<td>Private English Courses</td>
<td>72 %</td>
<td>84 %</td>
</tr>
<tr>
<td>High School English Only</td>
<td>27 %</td>
<td>15 %</td>
</tr>
<tr>
<td>Male</td>
<td>75 %</td>
<td>58 %</td>
</tr>
<tr>
<td>Female</td>
<td>25 %</td>
<td>42 %</td>
</tr>
<tr>
<td>Schools Abroad</td>
<td>0 %</td>
<td>4 %</td>
</tr>
<tr>
<td>Visited English-speaking Country</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-2 months</td>
<td>31 %</td>
<td>38 %</td>
</tr>
<tr>
<td>Lived English-speaking Country</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-5 years</td>
<td>4 %</td>
<td>2 %</td>
</tr>
<tr>
<td>Experience in Computer Science</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>59 %</td>
<td>55 %</td>
</tr>
<tr>
<td>Little</td>
<td>15 %</td>
<td>16 %</td>
</tr>
<tr>
<td>Large</td>
<td>25 %</td>
<td>26 %</td>
</tr>
<tr>
<td>No answer</td>
<td>1 %</td>
<td>3 %</td>
</tr>
</tbody>
</table>

**Materials**

The placement test administered to all students is general in purpose, and students are required to take it before enrolling for the first
time in the computer science program. Students' scores are evaluated according to the structure of our program and they are designated to enroll in the course that best matches their level of performance. As a consequence of having administered this test for many years, our classes have been organized into fairly homogeneous groups (Becher & Oliveira, 1989) enabling both learners and teachers to better profit from the learning/teaching experience.

The use of this type of test for so many years has also engendered continuous changing of the instrument so that it has finally become a reliable instrument that places students according to the different levels of language maturity that they have when they begin to study English at PUC-Rio.

In general terms, the placement test aims at measuring students' understanding of English as used for a special purpose. More specifically, the goal of the test is to evaluate the extent to which students can read the language used in computer science. To some degree, the test questions are designed to differentiate content knowledge from language knowledge. The test separates students into homogeneous groups that will be taught similar language points based on both language knowledge and language skills. The students are placed into one of the courses required (Level I or Level IA).

The original version of the placement test (see Appendix A) was composed of two parts. Part I, based on an authentic 450-word article extracted from a computer magazine, consisted of eight productive test items. This section of the test was designed to measure the students' abilities to use strategies considered crucial to the reading process (skimming and scanning for information and vocabulary, interpreting the author's intention and point of view). Part II, based on two 10-12 line paragraphs extracted from a computer science textbook and a brief letter published in a computer magazine, consisted of 25 receptive test items. It was designed to assess the students' knowledge of technical vocabulary and connectives in context. Considerably more detail will be provided about this test in the Results section of the paper.

Procedures

The original and revised versions of the test were administered under similar conditions in May and October 1992. On both occasions, the same auditorium was reserved for the computer science students.
to take the test. The maximum time allowed was approximately two hours. In October, however, the subjects taking the revised version were also asked to complete a 40-item cloze test within 25 minutes. Once they had completed the cloze test, they proceeded on with the placement test.

**Analysis**

The original test and the revised version were analyzed with the help of a software program called QuattroPro™ (1991) on an IBM compatible personal computer. First, all of the data were entered into this spreadsheet program. Each students' test was given an identification number which was marked on the test paper itself, then this number was entered into the first column of the spreadsheet. Each students' answers were then entered into subsequent columns as 1's (for correct) and 0's (for incorrect) such that each row of the spreadsheet represented on student's performance on the test – item by item. The total scores (the sum of all the ones and zeros) were then calculated and put into the last column. Next, the student records (rows) were sorted so that the total scores were arranged from the highest score to the lowest one in consistently descending order.

**Item analysis**

Each individual item on the test was then analyzed according to the following item analysis statistics:

*Item facility* (IF) is a statistical index used to examine the percentage of students who correctly answered a given item. It is calculated by adding up the number of students who responded correctly to a question and dividing that sum by the total number of students taking the test. IF statistics range from 0.00 for items that no student answered correctly to 1.00 for items that all students answered correctly, and, of course, it can take on all the values in between. The following (Garrett) scale may help in assessing the IF statistics:

- **00 – .15** Very difficult
- **16 – .50** Difficult
- **51 – .85** Easy
- **86 – 1.00** Very easy

138  
Alfa, São Paulo, 41: 133-158, 1997
Item discrimination (ID) is an indicator of the degree to which an item separates the "high achievers" (with high scores) from the "low achievers" (with low scores). Here, it was calculated by contrasting the performance of the upper third of the students with that for the lower third. The ID for each item was calculated by first computing the item facility for the high achievers and low achievers, separately, then subtracting the IF for the low group from the IF for the high group. ID can range from .00 to 1.00 (the higher the value the better because higher values indicate items that are discriminating well between the two groups of students) and from .00 to -.1.00 (the higher the negative values indicate items that are acting in some way different from the whole test). The following scale may help in interpreting the ID statistics:

<table>
<thead>
<tr>
<th>ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00 - .20</td>
<td>Very low (discrimination)</td>
</tr>
<tr>
<td>0.21 - .26</td>
<td>Poor</td>
</tr>
<tr>
<td>0.27 - .33</td>
<td>Average</td>
</tr>
<tr>
<td>0.34 - .40</td>
<td>Good</td>
</tr>
<tr>
<td>0.41 - .60</td>
<td>Very good</td>
</tr>
<tr>
<td>0.60 - 1.00</td>
<td>Excellent</td>
</tr>
</tbody>
</table>

The two statistics taken together can be used to select those items that are functioning well for placement decisions for the particular group in question. Following Brown (1992) and Ebel (1979), those items which have overall IF values between .30 and .70 and relatively high ID values are items that should be retained in the revised version of the test, and those items that have overall IF values below .29 or above .71 and low ID's are items that should be deleted from the test.

Descriptive statistics

Descriptive statistics are numerical representations describing how the groups performed on the test. The following descriptive statistics were used in this study:

The Mean (X) is one indicator of the central tendency, or typical performance of the group. It is essentially the same as the arithmetic average of the scores. The mean was calculated by adding up all of the scores and dividing the result by the total number of scores.
The **Standard deviation** is a sort of average of the differences of all scores from the mean. It is an indicator of the dispersion of scores around the mean.

**Reliability** coefficients, as used here, indicate the degree to which a test is internally consistent, or reliable. The Kuder-Richardson formulas 20 and 21 were used in this study (K-R20 and K-R21). These coefficients can range from .00 to 1.00 and, by moving the decimal point two places to the right, can be directly interpreted as the percent of consistent variation in the scores on a test. Thus a test with a reliability coefficient of .85 can be said to be 85% reliable.

**Validity** is defined here as the degree to which a test measures what it claims to be measuring, and it can only be examined after the reliability of the test is determined to be acceptable. Two strategies will be used here. The first is the criterion-related validity strategy, which involves examination of the correlation of the scores on a test with some outside measure (in this case, cloze test scores). The second strategy used here in thinking about the validity of the test is called content validity, i.e., the validity of this test was defended on the basis of the argument that the content of the test is a representative sample of the types of English language that computer science students will need in taking their courses and pursuing their careers in that field.

**Results**

The results of this paper will be presented in an order that corresponds to the order of the research questions given in the introduction to the paper. Thus the item analysis results will be presented first followed by the descriptive statistics, as well as reliability and validity statistics.

**Item Analysis**

This section of the paper will present a detailed description of the items in the original version of the placement test as well as a description of the revised version, highlighting only those items which were modified or substituted as a result of the item analyses carried out in improving the test. The item statistics for the original and revised versions of the test are presented in Table 2.
Table 2 - Item statistics for the original and revised versions

<table>
<thead>
<tr>
<th>Original item number</th>
<th>Original IF</th>
<th>Original ID</th>
<th>Revised item number</th>
<th>Revised IF</th>
<th>Revised ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>I1</td>
<td>.63</td>
<td>.67</td>
<td>I1</td>
<td>.60</td>
<td>.48</td>
</tr>
<tr>
<td>I2A</td>
<td>.42</td>
<td>.56</td>
<td>I2B*</td>
<td>.56</td>
<td>.15</td>
</tr>
<tr>
<td>I2B</td>
<td>.53</td>
<td>.78</td>
<td>I2C</td>
<td>.69</td>
<td>.68</td>
</tr>
<tr>
<td>I2C</td>
<td>.21</td>
<td>.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I2D</td>
<td>.58</td>
<td>.67</td>
<td>I2E</td>
<td>.69</td>
<td>.61</td>
</tr>
<tr>
<td>I2E</td>
<td>.32</td>
<td>.44</td>
<td>I2F*</td>
<td>.33</td>
<td>.17</td>
</tr>
<tr>
<td>I2F</td>
<td>.32</td>
<td>.22</td>
<td>I2G</td>
<td>.61</td>
<td>.45</td>
</tr>
<tr>
<td>I2G</td>
<td>.37</td>
<td>.56</td>
<td>I2H*</td>
<td>.35</td>
<td>.18</td>
</tr>
<tr>
<td>I2H</td>
<td>.26</td>
<td>.33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I3A</td>
<td>.37</td>
<td>.78</td>
<td>I3A</td>
<td>.29</td>
<td>.60</td>
</tr>
<tr>
<td>I3B</td>
<td>.32</td>
<td>.44</td>
<td>I3B</td>
<td>.17</td>
<td>.37</td>
</tr>
<tr>
<td>I3C</td>
<td>.26</td>
<td>.56</td>
<td>I3C</td>
<td>.26</td>
<td>.60</td>
</tr>
<tr>
<td>I3D</td>
<td>.26</td>
<td>.56</td>
<td>I3D</td>
<td>.29</td>
<td>.44</td>
</tr>
<tr>
<td>I3E</td>
<td>.32</td>
<td>.44</td>
<td>I3E</td>
<td>.43</td>
<td>.54</td>
</tr>
<tr>
<td>I3F</td>
<td>.21</td>
<td>.44</td>
<td>I3F</td>
<td>.29</td>
<td>.63</td>
</tr>
<tr>
<td>I4</td>
<td>.52</td>
<td>.78</td>
<td>I4</td>
<td>.45</td>
<td>.60</td>
</tr>
<tr>
<td>I5</td>
<td>.53</td>
<td>.78</td>
<td>I5</td>
<td>.43</td>
<td>.70</td>
</tr>
<tr>
<td>I6</td>
<td>.74</td>
<td>.56</td>
<td>I6A</td>
<td>.56</td>
<td>.70</td>
</tr>
<tr>
<td>I7</td>
<td>.68</td>
<td>.33</td>
<td>I7</td>
<td>.72</td>
<td>.58</td>
</tr>
<tr>
<td>I8A</td>
<td>.42</td>
<td>.78</td>
<td>I8A</td>
<td>.19</td>
<td>.40</td>
</tr>
<tr>
<td>I8B</td>
<td>.16</td>
<td>.33</td>
<td>I8B</td>
<td>.13</td>
<td>.30</td>
</tr>
<tr>
<td>I8C</td>
<td>.05</td>
<td>.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I8D</td>
<td>.00</td>
<td>.00</td>
<td>I8C</td>
<td>.19</td>
<td>.40</td>
</tr>
<tr>
<td>I8E</td>
<td>.26</td>
<td>.44</td>
<td>I8D</td>
<td>.18</td>
<td>.37</td>
</tr>
<tr>
<td>IIA1</td>
<td>.79</td>
<td>.22</td>
<td>II1</td>
<td>.40</td>
<td>.34</td>
</tr>
<tr>
<td>II1A2</td>
<td>.42</td>
<td>.78</td>
<td>II2</td>
<td>.51</td>
<td>.54</td>
</tr>
</tbody>
</table>

Alfa, São Paulo, 41: 133-158, 1997
<table>
<thead>
<tr>
<th>Original item number</th>
<th>Original</th>
<th>Revised item number</th>
<th>Revised</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IF</td>
<td>ID</td>
<td></td>
</tr>
<tr>
<td>II1A3</td>
<td>.47</td>
<td>.67</td>
<td></td>
</tr>
<tr>
<td>II1A4</td>
<td>.53</td>
<td>.78</td>
<td>II4</td>
</tr>
<tr>
<td>II1A5</td>
<td>.63</td>
<td>.11</td>
<td>II5</td>
</tr>
<tr>
<td>II1A6</td>
<td>.63</td>
<td>.56</td>
<td>II6</td>
</tr>
<tr>
<td>II1A7</td>
<td>.53</td>
<td>.56</td>
<td>II7*</td>
</tr>
<tr>
<td>II1A8</td>
<td>.42</td>
<td>.33</td>
<td>II8</td>
</tr>
<tr>
<td>II1A9</td>
<td>.42</td>
<td>.22</td>
<td>III1*</td>
</tr>
<tr>
<td>II1A10</td>
<td>.32</td>
<td>.67</td>
<td>II9</td>
</tr>
<tr>
<td>II1B11</td>
<td>.68</td>
<td>.33</td>
<td>III2*</td>
</tr>
<tr>
<td>II1B12</td>
<td>.74</td>
<td>.56</td>
<td>III3</td>
</tr>
<tr>
<td>II1B13</td>
<td>.53</td>
<td>.67</td>
<td>III4</td>
</tr>
<tr>
<td>II1B14</td>
<td>.16</td>
<td>.11</td>
<td>III5</td>
</tr>
<tr>
<td>II1B15</td>
<td>.47</td>
<td>1.00</td>
<td>II10</td>
</tr>
<tr>
<td>II1B16</td>
<td>.32</td>
<td>.44</td>
<td>II12</td>
</tr>
<tr>
<td>II1B17</td>
<td>.21</td>
<td>.22</td>
<td></td>
</tr>
<tr>
<td>II1B18</td>
<td>.68</td>
<td>.67</td>
<td></td>
</tr>
<tr>
<td>II1B19</td>
<td>.37</td>
<td>.56</td>
<td></td>
</tr>
<tr>
<td>II21</td>
<td>.05</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>II22</td>
<td>.26</td>
<td>-.22</td>
<td></td>
</tr>
<tr>
<td>II23</td>
<td>.21</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>II24</td>
<td>.26</td>
<td>.22</td>
<td></td>
</tr>
<tr>
<td>II25</td>
<td>.11</td>
<td>.00</td>
<td></td>
</tr>
</tbody>
</table>

* Items which had been selected for elimination in the projected version of the test to be administered in the following semester.

142  
Alfa, São Paulo, 41: 133-158, 1997
Notice that the original item numbers are presented in the first column of Table 2 with the corresponding IF and ID values in the second and third columns. The revised item number and corresponding IF and ID are presented in columns four to six. Items that were eliminated in the revised version of the test are blank in the last three columns, while items that were added to the revised version are blank in the first three columns. Those revised items with an asterisk have been selected for further elimination in the projected version of the test which was administered the next time it was used (in our ongoing process of test improvement).

It is important to understand that both versions of the test were made up of individual items and clusters of items that will be referred to as subtests in the following text. In accordance with the criteria stated in the Analysis section above, item I1, which assessed the students' abilities to interpret the author's declared intention, worked very well (IF = 63; ID = .67) and was therefore retained in the revised version of the test. The second and third subtests were similar in that they were made up of subsets of fill-in items which required the students to complete two tables using information contained in the text.

The second subtest was divided into items I2A to I2H. It was noticed, however, that this subtest should have been clearer in directions to the students and scorers, alike. The first two items in subtest 2, where price information was expected, were not counted in the total scores on the original test due to scoring inconsistencies. The rest of the items in subtest 2 (I2A to I2H) included five which were ideal for retention in the revised version of the test with IF values between .30 and .70 and ID values higher than .44. However, there were also three problematic items, which needed to be rejected or improved. Item I2C (IF = .21; ID = .00) and item I2H (IF = .26; ID = .33) were eliminated by filling both spaces in the table with the expected answer. This solution was also seen as a way of clarifying to the students what was expected of them. Clearer scoring instructions were also included in the answer key. Item I2F (IF = .32; ID = .22) was considered marginal due to its low ID but was retained because its IF was slightly above the .30 cut-point.

The third subtest contained items I3A to I3F. Although four items had IF values below .30, the ID values were relatively high, ranging from .44 to .78. Consequently, these items were retained in the revised version of the test.

Items 4 to 7 were short-response items in which students were expected to answer in one of two full sentences. It was quite clear that
the first three items in this sequence were good items that should be
retained in the revised version since their IF values were between 53
and 74 and their ID values were quite high (78, 78, and 56, respectively)
In the revised version, Item 6 was divided into two sub-items, I6A
and I6B for the sake of clarity in answering and in scoring Item 7,
which might have been subject to improvement due to its ID value of
33, was left intact so that there should be at least one easy item in this
series

In items I8A to I8E, the students were expected to find the text
synonyms for the words given in the items Unfortunately, however,
three out of the five items turned out to be problematic for an item
analysis point of view Item I8B (IF = 16, ID = 33) proved to be too dif-
cult but was included in the revised version because it had an accep-
table ID value Item I8C was definitely a poor item and was eliminated
in the revised version of the test The expected word in the text – cum-
bersome – was much too difficult for most students and the word that
could have been accepted as the second best answer did not attract
enough respondents for the answer key to be altered Item I8D presen-
ted a typographical error which had gone unnoticed during the test
development process This error may have led all of the students to
leave the item unanswered This item was corrected and retained in
the revised version of the test Although item I8E had an IF of 26, it was
kept because of its relatively high ID value of 44

The twenty items which make up the first subtest of Part II on the
test were receptive in nature since they required students to select
from a list of optional technical vocabulary in order to complete two
short paragraphs Twelve of these items worked quite well (with IF
values of 42 to 74 and high ID values) and were retained for the revised
version of the test (It is worth noting that item II1B15 was a perfect dis-
criminator ) On the other hand, there were a total of nine items in the
II1A subtest (items 1, 5, 8, 9, 11, 14, 16, 17, and 20) which needed to be
discarded because they had relatively low ID values The decision to
drop these nine items was endorsed by the group, who saw this deci-
sion as a justified way of shortening and cutting down on the number
of technical vocabulary items The texts in the revised version were
kept the same, but each had only six blanks that the students were
expected to fill in The five items (II21 to II25 which composed subtest
2 of Part II in the original version, set out to measure in a receptive-pro-
ductive manner the students’ abilities to use connectives These items
proved to be quite ineffective since their IF values all fell below .30 (ranging from .05 to .26) and the ID values were low or even negative. Apparently, this set of items was too difficult – in part, perhaps, because of formatting problems. Consequently, these items had to be deleted from the revised version of the test. However, content validity considerations required that this linguistic aspect of the language be tested. In addition, it was considered important that some combination of receptive-productive items be included in the test. As a result, a relatively easy passage was extracted from a computer science textbook and a whole new set of items was created.

Descriptive statistics

Table 3 gives the descriptive statistics for the original version of the test, for the revised version and for the version that was projected for the next administration. For the projected version, these statistics are the result of doing a reanalysis of the revised version results with the weak items (the ones with asterisks in column four of Table 2) eliminated.

Table 3 – Descriptive statistics for the original, revised and projected versions of the test

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Original</th>
<th>Revised</th>
<th>Projected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of items</td>
<td>49.00</td>
<td>42.00</td>
<td>36.00</td>
</tr>
<tr>
<td>Mean</td>
<td>19.63</td>
<td>19.60</td>
<td>16.83</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>10.94</td>
<td>8.97</td>
<td>8.37</td>
</tr>
<tr>
<td>Maximum</td>
<td>36.00</td>
<td>39.00</td>
<td>35.00</td>
</tr>
<tr>
<td>Minimum</td>
<td>3.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Range</td>
<td>34.00</td>
<td>40.00</td>
<td>36.00</td>
</tr>
<tr>
<td>Average IF</td>
<td>.40</td>
<td>.47</td>
<td>.47</td>
</tr>
<tr>
<td>Average ID</td>
<td>.44</td>
<td>.48</td>
<td>.53</td>
</tr>
</tbody>
</table>

Notice that the first row gives the total number of items for the three versions of the test, and that subsequent rows provide the mean, standard deviation, maximum score obtained, minimum score obtained, score range, average item facility and average item discrimination.
Reliability

The reliability statistics are presented in Table 4. Notice that the three versions of the test are approximately equal in reliability even though the revised and projected versions of the test become increasingly shorter. Since it is well known that shorter tests are generally less reliable than long ones, our test appears to be increasingly efficient in that it remains about equally reliable but is shorter in the revised and projected versions.

Table 4 – Reliability statistics for the original, revised, and projected versions of the test

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Original</th>
<th>Revised</th>
<th>Projected</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-R20</td>
<td>.94</td>
<td>.91</td>
<td>.92</td>
</tr>
<tr>
<td>K-R21</td>
<td>.92</td>
<td>.90</td>
<td>.90</td>
</tr>
</tbody>
</table>

Validity

Two strategies were used here. The first was the criterion-related validity strategy, which involved examination of the correlation of the scores on a test with some outside measure (in this case, a cloze test). The correlation between the revised version of our test and the cloze test scores .62 (n = 93), which lends support to the notion that our test is valid in the same sense that a cloze test is, i.e., for testing overall language proficiency.

The second strategy used here was content validity. We feel that the validity of our test can be defended on the basis of the argument that the content of the test was carefully set up to be a representative sample of the types of English language that computer science students will need in taking their courses and pursuing their careers in that field. Each item was reviewed by at least three of the English faculty with this criterion in mind when the test was originally created and in subsequent versions. Efforts were also made to balance the item types and content in about the same way in all versions of the test.
Discussion

1 How many test items in the original test need revision?

As shown in Table 3, the number of items consistently decreases as the test was reworked from the original to the revised and to the projected versions from 49 to 42 to 36 items, respectively. However, the answer to this question is not quite that simple. As indicated in Table 2, 18 items were eliminated from the original version in creating the revised version. Then, 10 new items were added to that revised version in order to maintain a balance of item types. A further six items (ones that did not function well) were removed in the projected version of the test.

2 To what degree is the score distribution of the original test improved in the revised version of the test?

As indicated in Table 3, the score distributions of the successive revisions of the test appear to become increasingly well-centered. In addition, there is room for at least two standard deviations above and below the mean on the revised and projected versions of the test (while this was not true for the original test).

3 To what degree is reliability affected by the revision process?

On the whole, all versions of the test were very reliable with KR20 and KR21 coefficients in excess of 90. However, as stated above, a shorter test that is equally reliable with a longer one is more efficient. Hence, our revision processes have improved the efficiency of our placement procedures by creating a shorter test that is approximately equal in reliability.

4 To what degree is the final version of the test valid?

The validity of the test was discussed from both the criterion-related and content validity points of view. Since the criterion-related validity results indicate that the test is assessing overall English language proficiency and since the content validity issues indicate careful planning in developing the computer science content, we feel confident that the test is measuring what we claim to be measuring to a satisfactory level.

Alfa, São Paulo, 41: 133-158, 1997
Conclusion

On the whole, this project has demonstrated the efficacy and importance of using item analysis techniques to revise language program tests. This process has not only given us information about individual items and allowed us to improve successive versions of the test, as well as to create and streamline new tests. In the past, there was a great deal of dissatisfaction with these placement procedures. However, given this on-going process of test revision and analysis, we believe we are now able to provide placement for our students that is not only consistent and accurate, but also fair.


- RESUMO: O presente artigo inicia-se com uma breve análise de conceitos básicos na área de testagem, apresentando, a seguir, a proposta dos testes de nivelamento aplicados aos alunos do curso de Tecnólogo em Processamento de Dados (TPD) da PUC-Rio - os quais devem, por exigência curricular, cursar um dos dois níveis de Inglês Instrumental para Informática oferecidos pelo Departamento de Letras. O estudo desenvolvido analisa os resultados de duas versões de um teste aplicado, em semestres consecutivos, a dois grupos diferentes de alunos de TPD. Em relação ao teste original, foram calculados os índices de facilidade e de discriminação, estimando-se a confiabilidade do teste. As estatísticas foram cuidadosamente analisadas, levando à eliminação, reformulação ou substituição de algumas questões, de modo a produzir-se uma nova versão do teste. Esta versão reformulada foi aplicada a um segundo grupo de estudantes, realizando-se o mesmo estudo estatístico. Os resultados ressaltam a importância da análise e reformulação das questões para garantir instrumentos de testagem mais consistentes e precisos.

- PALAVRAS-CHAVE: Testagem; ensino; língua inglesa.

References


DEPARTAMENTO DE LETRAS
PUC/RIO
PROVA DE NIVELAMENTO EM LÍNGUA INGLESA
MAIO 1992

Nome do aluno: .................................................................
Matrícula nº: .................. Departamento: ......................
Data: .../.../...

RESPONDA POR FAVOR

1. Quantos anos de inglês você já estudou? .................................
2. Qual foi o último nível cursado? ............................................
   Em que instituição? ................................................................
   Em que ano? ........................................................................
3. Você já morou ou visitou alguma pais de língua inglesa? ............
   Qual? .................................................................
   Quanto tempo? ...........................................................
4. Assinale na escala abaixo, o seu grau de conhecimento em inglês:
   0 1 2 3 4 5
   Compreensão escrita  ( ) ( ) ( ) ( ) ( ) ( )
   Compreensão oral ( ) ( ) ( ) ( ) ( ) ( )
   Produção oral ( ) ( ) ( ) ( ) ( ) ( )
5. Numere em ordem decrescente os itens em que você sente maior dificuldade ao ler um texto de área de Processamento de Dados em inglês:
   ( ) Estruturas gramaticais
   ( ) Vocabulário em geral
   ( ) Vocabulário técnico
   ( ) Conhecimento básico na área de informática.
   ......................................................................................
   ......................................................................................
   ......................................................................................

NÃO PREENCHER

Nívelado para:
Inglês Técnico I ( ) Inglês Técnico IA ( )

Assinatura do professor: ..............................................................

Alfa, São Paulo, 41: 133-158, 1997
If you’re serious about color painting take my advice and part with a few extra bucks to buy a full-fledged color photo-editing package. ColorStudio and Adobe Photoshop, the most powerful examples, sell for about $300 more than either PixelPaint Pro or Studio! 32. Although they lack tools for creating geometric shapes like rectangles and ovals, both are far superior products for many high-end color paint projects. Both Photoshop and ColorStudio provide 24-bit color using an 8-bit monitor; virtual memory; a slew of useful image-editing filters; full contrast, brightness, and color balance control; anti-aliasing; and pressure sensitivity. In addition, both offer exciting control when creating four-color separations, and they provide direct support for scanners.

Deciding between the two, however, is a difficult proposition for the discerning artist. ColorStudio offers an extremely versatile collection of drawing tools and brush shapes. The program also allows you to customize brush shapes, which can be saved and transferred to other versions of ColorStudio as annex files—special resources that can include separation modules and image-editing filters. From an electronic painter’s perspective, ColorStudio’s greatest strength over Photoshop may be the Shapes annex, a collection of object-oriented tools and commands that rival those of dedicated drawing programs such as Adobe Illustrator and Aldus FreeHand. Like Illustrator 3.0, ColorStudio’s Shapes allows you to reshape PostScript type outlines to create logos and other specialized letterforms.

Unfortunately, while it’s a dramatic improvement over its predecessor in many ways, ColorStudio 1.11 has inherited two frustrating ancestral traits. First, it demands 512MB of RAM to run—more than twice as much memory as Photoshop requires—preventing many users from operating ColorStudio under MultiFinder. And second, the program has a cumbersome user interface, cluttered with complicated dialog boxes. Photoshop, on the other hand, delivers high-powered painting prowess in a substantially simplified format. A highlight among Photoshop’s points of painting interest is the magic eraser tool, which selectively reverts portions of a painting to their previously saved appearance. ColorStudio lacks even a standard eraser. What’s more, Photoshop offers versatile transformation commands and straightforward selection and other specialized letterforms.

But regardless of which program you choose—ColorStudio or, my favorite, Photoshop—there’s no getting around the fact that these two image-editing applications double as the best color paint programs on the market.
Now, answer questions 1 through 8. You may answer questions 4 through 7 in Portuguese.

1. The text aims at ...
   ( ) showing the causes that led to the creation of a photo editor.
   ( ) comparing and contrasting photo editors.
   ( ) showing the process used when dealing with a photo editor.
   ( ) describing the advantages and disadvantages of using a computer.

2. Complete the table with information contained in the first paragraph.

<table>
<thead>
<tr>
<th></th>
<th>Price</th>
<th>Object-oriented drawing tools</th>
<th>Direct scanner support</th>
</tr>
</thead>
<tbody>
<tr>
<td>ColorStudio</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adobe Photoshop</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PixelPaint Pro</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Studio/32</td>
<td>$699</td>
<td></td>
<td>$699</td>
</tr>
</tbody>
</table>

3. In the 2nd and 3rd paragraphs ColorStudio and Photoshop have some of their features compared. Complete the table below with comments extracted from the text on some characteristics of the two programs.

<table>
<thead>
<tr>
<th>Shapes annex</th>
<th>Eraser</th>
<th>Transformation commands/ masking control</th>
</tr>
</thead>
<tbody>
<tr>
<td>ColorStudio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Photoshop</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. Which is the advantage of ColorStudio over Photoshop?

5. Which is the advantage of Photoshop over ColorStudio?

6. Mention two disadvantages of ColorStudio?

7. Which program does the author prefer?

8. Find in the 2nd or 3rd paragraphs of the text words that mean:
   a) change the form ..............................................................
   b) characteristics .............................................................
   c) complex,difficult ...........................................................
   d) complete with ...............................................................
   e) full of .................................................................................
PART II

1. Using some of the terms in the list below, complete the blanks in paragraphs (a) and (b).

(a) At the centre of the computer systems is the (1) which can be considered to be the 'brain' of the computer. Its main components are the central processor and the main (2). The speed and capacity of these (3) have been greatly improved with each new (4) of computers.

In the first generation, the central processor was built from electronic (5) which were rather unreliable. The second generation used (6). The third generation used integrated (7). The fourth generation of computers uses (8). These are contained on electronic chips which are slices of (9) with thousands of (10) components and circuits engraved on them.

(b) The operating system is the most important type of systems software. It consists of a group of (11) designed to manage and co-ordinate all the (12) and software of a computer system as efficiently as possible and to provide communications between the computer and the (13). It is a very complex piece of (14) which performs many different (15) such as controlling the operation of the disk (16) displaying prompts and (17) keeping the system running if an (18) occurs in a program, checking the input of identification (19) and passwords and keeping a log of terminals used in a multi-user system.

The operating system must be (20) with the central processor and is usually supplied by the computer manufacturer.

2. First read through the following text. Then, fill in the blanks with some of the connectives below:

because therefore despite likewise yet however but while

COLOR COMPARISON

We at Adobe want to comment on Ion Poole's recent comparison of Letraset ColorStud...
hour with Photoshop than creators of ColorStudio Mark Zimmer and Tom Hedges did in two hours and 20 minutes and than Letraset employee Tony DoYoung did in two hours with ColorStudio. (1) A clear explanation of the retouching techniques used by both teams, Poole credits operator skill as the sole reason for the superior results achieved by the Photoshop team (2) the products compared are in fact very similar.

(3) John Knoll is, admittedly, a competent retoucher, we think the article's conclusions ignore the very evidence this test was supposed to bring to light: which product performs best for this type of retouching work. (4) we think the printed results of this comparison speak for themselves. Unfortunately, (5) the article still leaves readers a bit confused and leaves it up to them to make their own decision.

Steve Guttman  
Mountain View, California

(1) ........................................ (2) ........................................
(3) ........................................ (4) ........................................
(5) ........................................
APPENDIX B

DEPARTAMENTO DE LETRAS
PUC/Rio
PROVA DE NIVELAMENTO EM LÍNGUA INGLESA
OUTUBRO 1992

Nome do aluno: ..............................................................
Matrícula no: ........................................ Department: ........................................
Data: ......../......../....

RESPONDA POR FAVOR

1. Quantos anos de inglês você já estudou? ........................................

2. Qual foi o último nível cursado? ........................................
   Em que instituição? ........................................
   Em que ano? ........................................

3. Você já morou ou visitou algum país de língua inglesa? ......................
   Qual? ........................................
   Quanto tempo? ........................................

4. Assinale na escala abaixo, o seu grau de conhecimento em inglês:

   0 1 2 3 4 5
   Compreensão escrita ( ) ( ) ( ) ( ) ( ) ( )
   Compreensão oral ( ) ( ) ( ) ( ) ( ) ( )
   Produção oral ( ) ( ) ( ) ( ) ( ) ( )

5. Numere em ordem decrescente os itens em que você sente maior dificuldade ao ler um texto da área de Processamento de Dados em inglês:
   ( ) Estruturas gramaticais
   ( ) Vocabulário em geral
   ( ) Vocabulário técnico
   ( ) Conhecimento básico na área de informática.


   ..............................................................
   ..............................................................
   ..............................................................

NÃO PREENCHER

Nivelado para:
Inglês Técnico I ( )  Inglês Técnico IA ( )
Assinatura do professor: ..............................................................
Now, answer question 1 through 8. You may answer questions 4 through 7 in Portuguese.

1. The text aims at ...
   ( ) showing the causes that led to the creation of a photo editor.
   ( ) comparing and contrasting photo editors.
   ( ) showing the process used when dealing with a photo editor.
   ( ) describing the advantages and disadvantages of using a computer.

2. Complete the table with information contained in the first paragraph.

<table>
<thead>
<tr>
<th></th>
<th>Price</th>
<th>Object-oriented drawing tools</th>
<th>Direct scanner support</th>
</tr>
</thead>
<tbody>
<tr>
<td>ColorStudio</td>
<td></td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Adobe Photoshop</td>
<td></td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>PixelPaint Pro</td>
<td>$ 699</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>Studio/32</td>
<td>$ 699</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

3. In the 2nd and 4th paragraphs ColorStudio and Photoshop have some of their features compared. Complete the table below with comments extracted from the text on some characteristics of the two programs. Don't give "yes" or "no" answers.

<table>
<thead>
<tr>
<th>Shapes annex</th>
<th>ColorStudio</th>
<th>Photoshop</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eraser</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transformation Commands/masking control</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. Which is the advantage of ColorStudio over Photoshop?

5. Which is the advantage of Photoshop over ColorStudio?

   a) ..............................................................
   b) ..............................................................

7. Which program does the author prefer?

8. Find in the 2nd or 3rd paragraphs of the text words that mean:
   a) change the form ..............................................................
   b) characteristics ..............................................................
   c) compete with ..............................................................
   d) full of ..............................................................
1. Using some of the terms in the list below, complete the blanks in paragraphs (a) and (b).

(a) At the centre of the computer (1) is the CPU which can be considered to be the 'brain' of the computer. Its main components are the central processor and the main (2). The speed and capacity of these components have been greatly improved with each new (3) of computers.

In the first generation, the central processor was built from electronic valves which were rather unreliable. The second generation used (4). The third generation used integrated (5). The fourth generation of computers uses microprocessors. These are contained in electronic chips which are slices of silicon with thousands of (6) components and circuits engraved on them.

(b) The operating system is the most important type of systems software. It consists of a group of programs designed to manage and co-ordinate all the hardware and software of a computer system as efficiently as possible, and to provide communications between the computer and the (7). It is a very complex piece of software which performs many different (8) such as controlling the operation of the disc drives, displaying prompts and cursors, keeping the system running if an (9) occurs in a program, checking the input of identification (10) and passwords and keeping a log of (11) used in a multi-user system.

Alfa, São Paulo, 41: 133-158, 1997
2. First read through the text below. Then, fill in each of the blanks with a different connective chosen from the list below:

A computer keyboard is similar to a typewriter keyboard because it has alphabetic and numeric keys in a QWERTY layout. However, there are some differences. For example, the computer keyboard has user-defined function keys. Their use depends on how they are programmed; (1), it varies from program to program. Some computer keyboards have a BREAK key. (2) it must be used with great care (3) it clears the program (4) all the data from the main memory.

(1) ........................................... (2) ...........................................
(3) ........................................... (4) ...........................................

but therefore
and for