Comparative Analysis of the EC1 and EC2 Comprehensive Exams for valuating the learning results in the College of Engineering of Universidad del Norte, Colombia

Camacho Amparo[1], Cortés Omar[2], Salas Jhonnathan[3]

[1,3]Engineering college, Universidad del Norte, Colombia
[2] Measurement and assessment center, Fundación Universitaria Konrad Lorenz, Colombia

acamacho@uninorte.edu.co[1]

Abstract

The central aim of this project was the design, validation and application of the EC1 and EC2 Comprehensive Tests for evaluating the learning results in students of the Engineering Programs at Universidad del Norte. The EC1 was applied to 213 students and it included the evaluation of the Education Basic Cycle with a total of four specific tests: Physics, Mathematics, Reading Comprehension and Chemistry; while the EC2 was applied to 288 students and included a set of six specific tests of the Professional Education Cycle for the Civil, Industrial, Electronics, Electric, Mechanical and Systems Engineering Programs. The design of the EC1 and EC2 Tests included the participation of a team conformed by professors of Universidad del Norte ; the structure of each of the tests was formulated according to the Educational Objective Taxonomy of Bloom to classify the course learning outcomes by area and/or curricular axes of each of the curricula and the items were revised by disciplinary and psychometric academic pairs. Results show a comparative vision of the item analysis, the reliability indicators, validity and the score analysis from the Classical Test Theory and the Item Response Theory with the use of SPSS and Winsteps Software.

Introduction

Within the frame of the Assessment Process developed in the Engineering College of Universidad del Norte since the adoption of the ABET 2000 Engineering Criteria, the comprehensive tests have been implemented for the evaluation of the Course Learning Outcomes in two moments during the curriculum development; the first one includes the education component of Science and Communication courses; the second one includes the discipline education in each engineering program. These tests, along with other follow up and knowledge evaluation tools provide the required information to measure the advancement of the Program Outcomes, the fulfilment of the program Educational Objectives and then, based on these results, to take decisions of curricular improvement in each engineering program.

I. Theoretical framework

The Classical Test Theory (CTT) has been historically the most widely known as a result of its simplicity and its intuitive method, but currently, the Item Response Theory (IRT) is enhanced since it overcomes some of the shortcomings found in CTT: • The reliability and validity conditions of the test depend on the participants sample to be tested [1]. • The estimation of the examinee abilities depends on the specific set of items included in the test. • The same number of points in a test may correspond to different patterns of response. • Among the main characteristics of IRT are the following: the reactive properties do not depend of the group from which they were obtained, the points describing the examinee abilities do not depend on the test as a whole, the model is expressed more at the reactive level than at the test level and it does not require parallel testing to determine the reliability index, offering a measure of the precision of each ability index [2].

This theory allows the location of the interactions between the examinees and the items in a continuum representing them, called Test Characteristic Curve. Therefore, the positions of subjects on the line will depend on their answers to the test reactives. At the same time, the items will be located in different positions depending on their difficulty [3], [4]. This way it is possible to evaluate and relate the ability levels of the subjects with the difficulty levels of the test. Other advantages of the IRT: There is the possibility to describe an item independently from the characteristics of the sample, the possibility to describe the feature level independently from the administered set of items and the possibility to predict the properties of a test, even before being applied [5].

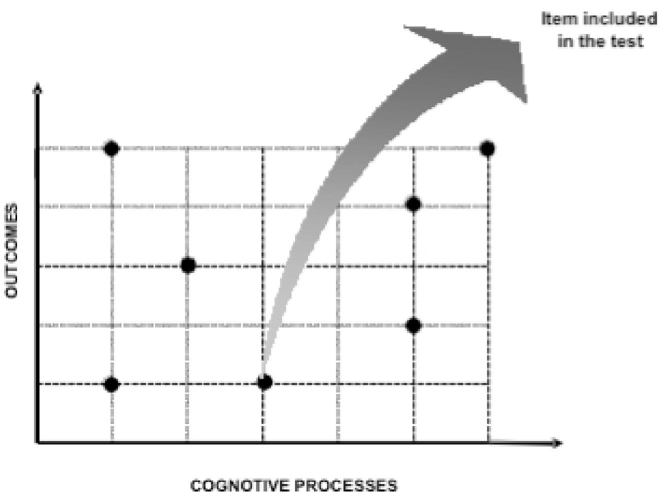
II. Test design and application

Method: A descriptive-comparative study of psychometric cut was developed where the differential behaviour of the EC1 and EC2 Comprehensive Test were analyzed from the models of the CTT and the IRT.
Participants: The sample was formed by 501 undergraduate students enrolled in the Engineering College in the basic (213) and professional education (288) cycles in the programs of Civil, Systems, Industrial, Electrical, Electronics, and Mechanical Engineering.
Instruments: Both tests were designed following the test specifications from Bloom Taxonomy in articulation with the different curricula in the Basic Education cycle for the EC1, as well as in the Professional Education cycle for the EC2. For the design and application of the comprehensive tests, three steps were followed:

a) Formulation of the EC1 and EC2 Comprehensive Tests specifications from Bloom Taxonomy, the Curricula Axes and the Course learning Outcomes. The test specifications comprise a detailed description of the instrument characteristics. The test structure was concentrated on a matrix representing the integration of the different dimensions of the evaluation objective to be measured, in the case of a discipline dimension and a dimension oriented to a thinking process [6]. The latter conceived from the point of view of Bloom Taxonomy, which divides cognitive processes in six categories: memory, comprehension, application, analysis, synthesis and evaluation [7].
Concerning the discipline dimension, curricular axes, sub-topics and course contents were taken into account. The curricular axes are the most general level within the curriculum, followed by the sub-topics and the course contents. Therefore, the evaluation of the outcome level of a course learning outcome is done from the evaluation of the course contents given the fact that they also include subtopics and topic axes.

Figure 1 shows the integration scheme of course learning outcomes and cognitive processes. The ultimate results of this integration are the questions applied in the test, in such a way that each confluence between a learning result and a cognitive process represents a question. It is important to say that in this sense, everything is within a curricular axis and that the scheme of Figure 1 is different for each sub-topic.

Figure 1. Integration scheme of learning results and cognitive processes



b) Design and Revision of Items with Expert Judgement Method. Experts' judgement is a procedure through which, a team of people, with expertise in the subject to be evaluated and in evaluation processes, evaluate the items of an instrument through the considerations of the conceptual foundation frame, in terms of its coherence with the evaluation aims and its pertinence with the test structure and the psychometric specifications and the characteristics of the target population; they also evaluate its degree of difficulty, its clarity and they offer suggestions for refining it [6].
c) Assembling the EC1 and the EC2 Tests in the Virtual Platform. In this step the tests were assembled in the Web on

the institutional portal of the university through the WebCT (Web Course Tools).

d) Administration of the EC1 and EC2 Comprehensive Test. The tests were applied by internet in the university campus and under the supervision of professors from the College. They have a total duration of six hours and are divided into two three-hour sessions.

e) Psychometric and Statistical Analysis of the EC1 and EC2 Comprehensive Tests with the support of the SPSS and Winsteps Programs. The psychometric analysis of the tests assumed a double level from the Test Classic Theory (CTT) and from the Item Response Theory (IRT) under the differential frame of Logistic Analysis from the Rasch Model with the support of the SPSS and Winsteps Programs.

III. Results Analysis

The results presented here correspond to those obtained in the first application of each test (EC1 and EC2).

Comprehensive Test I (EC1): the results obtained for EC1 in the frame of classical theory are illustrated in Figure 2 where they were compared with three score systems as follows:

1. Pure Performance: It corresponds to the estimated performance percentage between the numbers of correctly approached items per student over the total items of the test.

2. Performance curve: It corresponds to the percentage estimated for each student over the maximum scores observed in each program.

3. ECAES Scores (Quality of Higher Education State Exams): They correspond to the normalized score with an average of 100 points and a 10 points deviation in each program [8]. For this reason, the average is similar in all programs.

In general terms, starting with Figure 2, it is observed that the pure performances are in an estimated range of 46% and 49% in average, while the performance curves presented a tendency between 68% and 76% in average with a tendency to better performance levels obtained for the programs of Systems, Mechanical and Electronics Engineering. Figure 3 shows the test characteristic curve for EC1 from the Item Response Theory scope.

Figure 2. Comparison of Scores in (CTT) per Program (EC1).

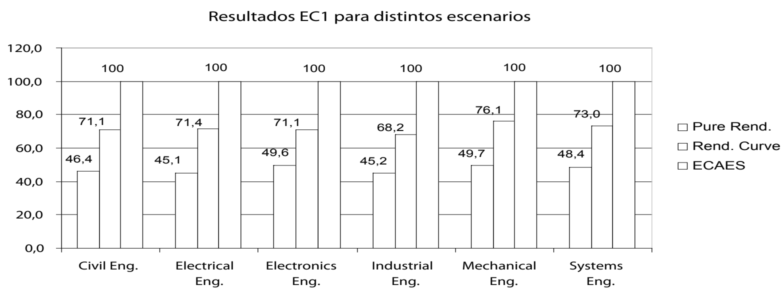
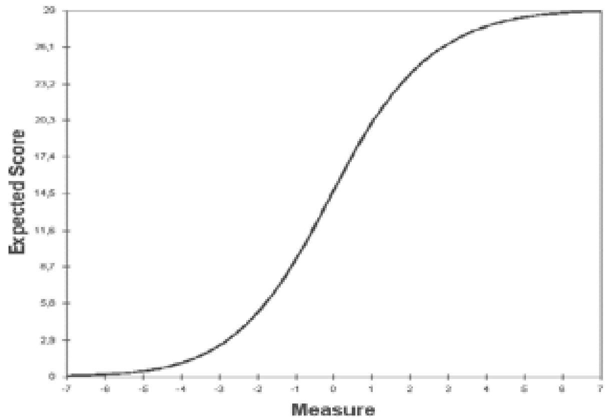


Figure 3. Comparison of Success (1) and Failure (0) curves of test (EC1)

```
         DICHOTOMOUS CURVES
P      ++ -------------- +-------------- +-------------- +-------------- ++
R  1.0 +                                                                 +
O      |                                                                |
B      |0                                                              1|
A      | 000000                                                         111111 |
B   .8 +        00000                                       11111       +
I      |             0000                              1111             |
L      |                 0000                      1111                 |
I      |                        000                111                     |
T   .6 +                        000          111                        +
Y      |                            000   111                           |
    .5 +                               ***                               +
O      |                            111   000                           |
F   .4 +                        111          000                        +
        |                     111               000                      |
R      |                 1111                        0000                    |
E      |             1111                          0000                 |
S   .2 +        11111                                       00000       +
P      | 111111                                                  000000 |
O      |1                                                                      0|
N      |                                                                |
S   .0 +                                                                +
E      ++ -------------- +-------------- +-------------- +---- ---------- ++
         -2               -1               0                1               2
          PERSON [MINUS] ITEM MEASURE
```

From the Rasch Model with the use of the Winstep Program the results derived previously are observed. There, the reliability index for the sample of 213 students was 0.72 while the reliability index for the group of 89 items was 0.98. This indicates that the test itself has a better level of consistency in contrast with the estimated from the CTT. The general estimation range for difficulty of the items was between -2,8 and 3,6 Logyt units(_), while the general estimation range for students ability was between -2,0 y 0,9 Logyt units (_). Figures 3 and 4 illustrate the General Test Characteristic Curve and the relation between the Success Probabilistic Curves (1) of ascendant cut and that of Failure (0) of descendent cut while the estimated ability level rises in Logyt units (_).

Figure 4. Test (EC1) General Test Characteristic Curve



Comprehensive Test II (EC2): Figure 5 shows the results of EC2 from CTT scope. It is observed that as in EC1 three score systems are compared and besides, that the pure performances range between 40% and 49% in average and the adjusted performance curves between 64% and 75% in average. In this sense it is important to say that tests are independent and therefore, scores correspond to specific domains of each program.

Figure 5. Comparison of Scores in (CTT) per Program (EC2).

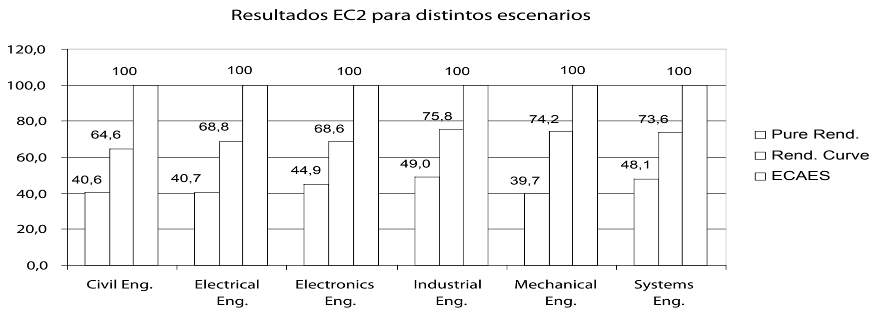


Table 1 presents a synthesis of the results obtained from the Rasch Model focus. One of the advantages of these results over the ones obtained in CTT is the possibility to estimate the reliability level in a differential way for the student response pattern and the psychometric behaviour of the test. In this sense the importance of analyzing the degree of consistency and convergence between the instrument and participants, with the expectancy that the reliability levels approach 1 in both cases

According to the parameters presented in Table 1 it can be noted that the psychometric behaviour concerning reliability level was good in most cases, having into account that the greatest divergence was observed in the Industrial Engineering Program. On the other hand, the student ability ranges presented the greatest variability levels in Electronics and Mechanical Engineering measured in Logyt (_) units. In the case of the difficulty levels the greatest variability in Logyt (_) units is observed in the Civil and Industrial Engineering Programs. Results show correspondence between the expected ability levels and the difficulty levels of the test.

Table 1. Synthesis of Psychometric Indicators from Rasch Model applied to EC2 Tests

	Civil Eng.	Electrical Eng.	Electronics Eng.	Industrial Eng.	Mechanical Eng.	Systems Eng.
Number of items	117	114	110	85	95	120
Number of students	19	33	55	105	19	57
Estimated media	32,2	32,9	41,4	23,9	23,32	50,72
Number of adjusted	82	80	91	48	55	100
Reliability levels of students Adjusted version	0,87	0,82	0,84	0,68	0,85	0,83
Reliability level of the test Adjusted version	0,78	0,83	0,92	0,96	0,79	0,93
General rank (Estimation of the ability level of students ?)	-1,5 a 1,5	-1,8 a 1,3	-1 a 1	-1 a 1,5	-1 a 2	-1 a 1
General rank (Estimation of the test difficulty level ?)	-3,5 a 3	-2 a 2,5	-2,5 a 2	-1 a 1,5	-3 a 3	-3 a 2,5

CONCLUSIONS

The implementation of the EC1 and EC2 tests has enriched learning in the item formulation and the test design administered through the use of virtual platforms. Strengths and weaknesses have been identified in the different areas of knowledge from the results point of view, as well as from the psychometric perspective of the student performance and the items estimated parameters. The most important result is the strengthening of the assessment process of the college, concerning the improvement of tests as measuring and essential support tools for the decision making processes oriented to the implementation of curriculum improvement actions.

References

01. Abad, J Temavi: Introducción a la Teoría de respuesta al Ítem. Universidad Autónoma de Madrid. Available: http://www.uam.es/personal_pdi/psicologia/fjabad/doctorado/TRI4_v2.pdf. (2006)
02. Hernández, A., Morales, V. y Maíz, J. La Teoría de Respuesta al Item (TRI) en la construcción de cuestionarios en Psicología del Deporte. Revista Digital. Año 10. No. 80. Available: http://www.efdeportes.com. (2005)
03. Muñiz, J. Introducción a la Teoría de Respuesta a los Ítems. Ediciones Pirámide: Madrid (1997)
04. Gregory, R. Evaluación Psicológica. Historia, principios y aplicaciones. 3ed. Manual Moderno: Bogotá D.C. (2001)
05. Olea, J., Ponsoda, V., Revuelta, J. y Belchi, J. Propiedades Psicométricas de un Test Adaptativo Informatizado de Vocabulario Inglés. Estudios de Psicología. 55, pp.61 – 73. (1996)
06. Rocha, M. y Pardo, C. Diseño de Pruebas de Evaluación Educativa. ICFES, Bogotá, Colombia (2005)
07. Herrera, A. Algunas consideraciones técnicas sobre la construcción de ítems de pruebas objetivas según la clasificación de objetivos educativos de Bloom. Memorias del Proyecto ECAES en Psicología. Bogotá: ICFES – ASCOFAPSI, (2003)
08. ICFES Estándares para la construcción de Pruebas, Bogotá, Colombia. (2004)