Educational Research (ISSN: 2141-5161) Vol. 3(5) pp. 457-467, May 2012
Available online@ http://www.interesjournals.org/ER

Full Length Research Paper

Psychometric analysis in knowledge-evaluation questionnaires, identification and implication

Gómez-Soberon J. Manuel*[1], Cabrera-Covarrubias F. Guadalupe[2], Arredondo-Rea S. Paola[3], Corral-Higuera Ramón[4], Almaral-Sánchez J. Luis[5], Gómez-Soberon M. Consolación[6], Gómez-Soberon U. José[7] and Pinedo Agustín[8]

[1*]Department of Architectural II. Universidad Politécnica de Cataluña, Spain
[2]Universidad Politécnica de Cataluña, Spain
[3/4/5]Universidad Autónoma de Sinaloa, Mexico
[6]Universidad Autónoma Metropolitana. Mexico
[7]Centro Escolar Presidente Francisco I. Madero, Mexico
[8]Universidad Nacional de la Plata, Argentina

Abstract

The experience achieved using the tool "Questionnaires", available inside the Virtual Campus of architectural engineering school in northeast Spain, is presented. "Questionnaires" is an adequate and simple instrument to evaluate the knowledge level achieved by students. This work shows and identifies the control indices of adaptation for the questionnaires, like the Facility Index, the Standard Deviation, the Discrimination Index and the Discrimination Coefficient. Derived from these parameters, the educational performances are inferred, identified and predicted. The conclusions of this work, permit to modify deficient knowledge-evaluation practices, to identify needs for specific groups or for students with particular requirements; being, in this way, feasible to apply these parameters with guarantee of success in similar evaluation processes.

Keywords: Educational adaptation, questionnaires in teaching, evaluation in teaching.

INTRODUCTION

Nowadays, conceptual and structural changes inside the European Educational Space (EEES) are being prompted in European universities, about which should be the work of professors, the form in how the knowledge should be transmitted, the easier way to learn for students, and finally, to achieve an accurate satisfaction according to the social context of a competent education required by the society and the educational institution (Veiga, 2009). In order to adhere to these objectives, in recent years, the High Education Institutions in Spain have started up improvement processes in their teaching practices (Gómez-Soberón, 2007) (Gómez-Soberón, et al., 2007) (Gómez-Soberón, 2009) (Gómez-Soberón, et al., 2009).

There are several published researches with reference to the application of formative evaluation methods in teaching, during the teaching process, having the objective to accomplish learning improvements. Introducing evaluation questionnaires, by means of data processing systems, could be a convenience and efficient strategy to reinforce the learning for students; some of the advantages of these systems are: the management of results, the velocity in the evaluation time, to avoid the use of paper to accomplish the questionnaires, etc. (Shepard, 2006). Nevertheless, there are also some objections, among others, the confidentiality of the student identity, the subsequent use of this information and its possible repercussion in educational process, etc. (Burden, 2008; Nulty, 2008); as well as the implications in be used as criteria only of information of the learning process of the students (Garfield, 2003).

There are few bibliographical references with reference to the analysis of the information generated with test-type questionnaires evaluations, which permits to professors having an idea of these results meaning,

*Corresponding Author E-mail: josemanuel.gomez@upc.edu

Table 1. Students in the course of study

Year	Course	Registered students
2002/2003	1Q	179
	2Q	209
2003/2004	1Q	339
	2Q	357
2004/2005	1Q	367
	2Q	374
2006/2007	1Q	367
	2Q	313
2007/2008	1Q	239
	2Q	290
2008/2009	1Q	252

and how they could became useful for professors and students (Romero et al., 2008; Blanco et al., 2009; Blanco, 2009; Blanco, 2010; Gómez-Soberón, 2010).

For the design of the evaluation tests, the multiple-option questions writing is a specialized labor that requires personnel with experience and training, thus, if they are adequate elaborated, they will permit to measure educational complex abilities (depending on the knowledge and experience of whom writes them). The description of items or questions implies to verify the relation between the item and the content supposedly measured by it; this verification is considered as a central part for the test validation processes (it is usual to verify it by student-professor feedback). Therefore, to confirm the validity of the test, the information about the content and the clarity and comprehension of items are important.

To achieve the previous statements, it is necessary to perform a statistical analysis to determine characterization indices of items (difficulty, correlation item-total score, discrimination and answer frequency according to option) and to select the adequate items for evaluation by theory of tests referred to norms, establishing as ideal questions those close to the 50% of difficulty and a discrimination above 0.40, and also, a correlation item- total score positive and significantly higher than zero (Esquivel, 2000).

On the other hand, note that both processes: the analysis of test results and application of questionnaires require extra effort for teachers, causing lost time from teaching or assessment. The main causes of lack of interest would be:

1. The refusal to new evaluation tools by the traditional professor.
2. The denial to investigate in situations that will not be repeated again.
3. “Extra” time in educational tasks.
4. Possibility to generate additional resources that cannot be assumed.

In contrast to the previous statements, it is important to note that currently there are tools and calculation processes that permit us the analysis of multiple processes, the generation of simulations or the validation of prediction hypothesis about guidelines inside the education field (Hutchison, 2009).

Within outlined context, in which the principal challenge for the educational development needs to generate mechanisms or evaluation systems that produce relevant information on what is taught and is learnt in a way effective in the schools, we have begun to deliberate all our educational processes. For this, we have gone incorporating the statistic use of the questionnaires that permits us to define parametric indices about the learning results of the students, as a new educational tool that gives response to the current demand of the analysis of the learning and deduction of possible tendencies or undesirable deviations in him.

Educational up to date framework and subject in study background

Constructions of concrete it is the course studied in this work; this course belongs to studies of university technical degree. This course is a four-month course in the second year of study (Obligatory in the Curriculum Block). This takes place in the four-month term 2B and it consists of six credits (not ECTS credits: European Credits Transfer System); subdivided in: 4.5 theoretical credits and 1.5 practices credits.

The subject is simultaneously given to four groups in all the four-month terms (1Q: Autumn and 2Q: Spring): two groups for students in the mornings (Group 1M and 2M) and two for students in the afternoons (3T and 4T).

The number of students that have studied this course in recent years is presented in Table 1. It is clearly observed that the number of students registered has been increased with the time, resulting in educational problems, such as: extra dedication by professors, decrease in teaching quality, difficulty to evaluate using traditional systems, etc.

It can be said, that in this discipline exists a high number of students that fail (for the most part repeating);

Table 2. Evaluation procedure of the subject

Thematic content (Modules)	Contents (%)	Technical of evaluation	Final Score (%)
1, 2, 3, 4 , 5 6, 7andy 8	50	Practical work No. 1	7.5
		1st. partial examination	40
9, 10 , 11 and 12	50	Practical work No. 2	7.5
9	15	Test 1*	5
10	15	Test 2*	5
9, 10 , 11 and 12	50	2nd partial examination	45

* Optional, not score in the final grade.

these is, that the students that approve the discipline are mostly those who are registered for the first time in it. On the other hand, the obtained scores are relatively low (Escuela Politécnica Superior de Edificación de Barcelona, 2009).

METHODOLOGY FOR DESIGN OF THE RESEARCH

For the design of the evaluation analysis system and to obtain the control indices to utilize in this work, some general criteria and practical recommendations have been followed to guarantee a correct application of the work and to avoid bias with an incorrect use of it (Ravela, 2000; Tiana, 1997). The information to be processed, answers calls response statistics for specific items, that is, used, analyzed and provide the percentages of multiple-choice items or percentages of student scores.

The starting point in the analysis process of the concerned data, the border of sampling was delimited to the course and group submitted for analysis, considering the following aspects:

1. Analysis period: 2008/09.
2. Course: 1Q.
3. Group where the analysis is done: 4T.

The motives by those which are reduced the analysis of this work to the previous variable, answers to dedication and adjustment motives of the courses imparted in the school; as well as by be a pilot experience, and by so much, defined for initial practice, calibration and verify of their suitability. The study group represents the total number of students in this group and course, and therefore the stratified response to a type of non-parametric sampling.

The analysis presented, obtains data to process from the results extracted of specific evaluations (two mid-term exams), from two works done by students and from two test (multiple option and test paired type) (Tuparov, 2008). The assessments of previous techniques are shown in Table 2.

The specific evaluations were individual; they consisted of solving graphic-conceptual problems. The activities developed by the students involved the resolutions of real cases with applications related to topics developed inside the classrooms. These activities were developed individually and valued according to some pre-established principles (rubric).

Test number 1 (multiple option) consisted of a thirty-tests with between three and five possible answers to select; while test number 2 (paired) consisted of four blocks of questions, in which each block contained among 8 to 12 questions, a total group of 41 questions; the structure of the two test assumes the implications and reasoning presented in the literature to this respect (Berrios, 2005). Both tests were implemented in the Virtual Campus of the course through the data processing platform Moodle (Dougiamas); though currently it is feasible to apply them in other similar platforms (Tuparov, 2008). This platform allows evaluations to take place virtually inside (our case) or outside the classroom, and evaluation is done by the previous test program, as a result of the process, the system generates an output file type Word, Excel or RTF allowing processing.

Tests were defined following next criteria and data processing adjustments, which help to standardize their application (these specifications were notified to the student body previous to their resolution):

1. Maximum time of resolution: 1 minute for each question.
2. Score considered as valid in each question: The last answer done.
3. Number of intents in one question: Unlimited.
4. Penalty in each question: The proportional part of the question value divided by the number of possible answers.
5. Value of each question: All the questions have the same proportional weight inside the global test.

The tests were also proposed to evaluate the different knowledge levels acquired by students, basing on the Taxonomy of Bloom (Van Niekerk, 2009). Table 3 summarizes the subdivision of knowledge levels evaluated, including the number of questions for each one of them.

Therefore, to analyze the first part of the statistical study, four different variables were used, to which, the codes and meanings presented in Table 4 were assigned. With the previous criteria and variables to analyze, the data processing program of statistical analysis SPSS V17 for Windows, was utilized, with the purpose of obtaining

Table 3. Bloom's Taxonomy of evaluative test

Level of knowledge	Subcategory	No. of questions, Test 1	No. of questions, Test 2
1	Knowledge	1	4
2	Understanding	9	13
3	Application	7	5
4	Analysis	9	9
5	Synthesis	1	0
6	Evaluation	3	10
	TOTAL	30	41

Table 4. Nomenclature of the study variables

Nomenclature	Meaning associated with the variable	Range of possible values
VAR01	Group to which belong the students	1M = 1, 2M = 2, 3T = 3 and 4T = 4
VAR02	Test 1	From 0 to 10*
VAR03	Test 2	From 0 to 10*
VAR04	Final Score	From 0 to 10*

* With accuracy of two decimal places of significance.

general descriptive statistical parameters of each variable, in a separated form, and thus to understand and distinguish them. The studied parameters were:

1. Central tendency measures (Mean, Median, Mode and Sum).
2. Dispersion measures (Standard deviation, Variance, Amplitude, Minimum, Maximum and Error of mean), Sampling distribution (Asymmetry and Kurtosis), and finally the percentile values.

The general results obtained for the four analyzed variables about theirs general statistical description, are presented in Table 5. Because of the amount of information developed in this article, only some interesting aspects of previous table are commented, remaining for the reader the total analysis of it following the recommendations set in the literature (Gómez-Soberón, 2010):

1. The mean qualification of students, which take the test, is higher than that of students which do not take it.
2. The variance of the results is smaller for the group that takes the test.

Psychometric analysis of items

Psychometric analysis is a mathematical procedure which applies statistical principles for determining the suitability of the proposed questions based on the responses and their individual relationship with the rest of the answers, thereby detecting whether the proposed questions are appropriate to assess the level of knowledge, degree of difficulty and degree of discrimination between high and low conceptual skills (Heck, 2006; Revuelta et al., 2003).

From the results of multiple-option and paired test, previously commented, some parameters have been extracted and utilized. They are defined and analyzed in Tables 6 and 7 where the processed data of the surveys are presented in a manner that permits the analysis and evaluation of the performance for each one questions taking into account the global evaluation of the sample. The statistical parameters utilized in this table have been determined with the evaluation of the classical theory of tests (Batrinca; General Public License GNU, 2010). For reasons of space in this paper only presents the first five questions and their assessments of the Multiple Choice Test, and the first two for the Test Paired.

The first parameter presented in the previous tables, for the analysis of the tests, is the Facility Index (FI, % correct), which is defined as the mean value of how easy or difficult an item is, with regard to the rest of questions inside the same analysis group (test). This parameter is determined with the following equation (1):

$$FI = X_{mean} / X_{max.} \quad (1)$$

Where:

X_{mean}, is the mean value from all values obtained for the total users who did every item.

$X_{max.}$, is the maximum value obtained for that item.

If the questions could be distributed in dichotomy categories (correct/incorrect), this parameter would coincide with the percentage of students that respond to the questions correctly.

In our study, and observing Figure 1; for the test 1, most of the questions are concentrated on the band from 70% to 90% of FI, while for the test 2 they are located in a band from 85% to 90%. From these results, it is deduced that the questions or blocks of questions located

Table 5. Descriptive statistics for study variables

Group	1M, 2M and 3T		4T					
			Test 1			Test 2		
Parameter	**VAR01**	**VAR04**	**VAR01**	**VAR02**	**VAR04**	**VAR01**	**VAR03**	**VAR04**
No. Valid*	247,00	247,00	32,00	32,00	32,00	39,00	39,00	39,00
No. Lost*	0,00	0,00	215,00	215,00	215,00	208,00	208,00	208,00
Mean	2,55	5,74	4,00	7,86	6,81	4,00	8,95	6,57
Standard error of the mean	0,07	0,14	0,00	0,18	0,26	0,00	0,29	0,22
Median	3,00	6,30	4,00	7,81	7,00	4,00	10,00	7,00
Mode	4,00	0,00	4,00	7,80	7,00	4,00	10,00	7,00
Standard deviation	1,13	2,27	0,00	1,04	1,45	0,00	1,80	1,38
Variance	1,28	5,17	0,00	1,08	2,09	0,00	3,24	1,91
Asymmetry	-0,06	-1,53		-0,86	-3,57		-2,44	-2,91
Standard error of Kurtosis	0,15	0,15	0,41	0,41	0,41	0,38	0,38	0,38
Kurtosis	-1,39	1,57		0,25	16,20		6,10	13,03
Standard error of Kurtosis	0,31	0,31	0,81	0,81	0,81	0,74	0,74	0,74
Amplitude	3,00	8,68	0,00	3,75	8,50	0,00	7,50	8,50
Minimum	1,00	0,00	4,00	5,39	0,00	4,00	2,50	0,00
Maximum	4,00	8,68	4,00	9,14	8,50	4,00	10,00	8,50
Sum	631,00	1.417,39	128,00	251,40	217,80	156,00	349,09	256,30
Percentiles 0	0,00	0,00	4,00	0,00	0,00	4,00	0,00	0,00
10	1,00	0,00	4,00	6,17	5,30	4,00	7,00	5,20
20	1,00	5,00	4,00	7,21	6,88	4,00	8,00	6,00
30	2,00	5,44	4,00	7,38	7,00	4,00	9,00	6,10
40	2,00	5,95	4,00	7,72	7,00	4,00	9,50	7,00
50	3,00	6,30	4,00	7,81	7,00	4,00	10,00	7,00
60	3,00	6,87	4,00	8,34	7,00	4,00	10,00	7,00
70	3,00	7,00	4,00	8,64	7,01	4,00	10,00	7,00
80	4,00	7,48	4,00	8,81	7,50	4,00	10,00	7,10
90	4,00	7,80	4,00	9,09	8,00	4,00	10,00	8,00
100	4,00	8,68	4,00	9,14	8,50	4,00	10,00	8,50

* Number of students included in each analysis.

out of both extremes of previous bands, should be eliminated in future editions of the test; since they are trivial (FI very low) or they are of a high difficulty level (FI very high). In any of both possibilities, these questions should not be utilized as criterion to discern an educational evaluation, since these questions are not useful as an evaluation criterion. The graph shows the areas discussed.

Another possible alternative for to decide which questions or blocks of questions could be eliminated from a test; is to verify that they be correctly defined, not including errors in their formulation and complying with basic criteria of logic. For which, an exhaustive review of its editing, structure, logic and coherence, must be done before using them again in another evaluation.

The second parameter evaluated in this work is the Standard Deviation (SD), in this parameter is studied the dispersion of the response in relation with the answers emitted by all the population analyzed. As a comment to this parameter, it can be said that in the event that all students respond equally to a specific question (item), the value obtained for the DS would be zero.

This parameter (SD) is obtained with the statistics standard deviation of the sample (classical analytic statistical); or if not, with the mark of class (relation obtained/maximum) for each specific item.

In our case, and observing Figure 2, this parameter can be utilized as criterion of detection to verify the knowledge acquisition by part of the student body, in a determined concept or item. This knowledge, contributed by SD, should not be seen in a way particular or individual, the correct interpretation is from a perspective most general and uniform of all the members (collective general knowledge of the theme).

In test 1, the questions that surpass the upper band of established criterion (for this case, it could be set in a SD close to 0.30), are questions with thematic content advisable to be reviewed again in the classroom, in order to guaranteeing some minimum contents assumed by all students.

Table 6. Details of test parameters for the study. Multiple Choice Test, Topic 9

No. Question	Possible answers	Possible value for each individual answers	Nº of times responded /Total Nº responded. For questions	Percentage of response answered by question	Nº Answers	FI (%)	SD	DI	DC
	1	1,00	30/32	94					
1	2	-0,33	0/32	0	32	82	0,336	0,958	0,775
	3		0/32	0					
	1	-0,33	0/32	0					
2	2		0/32	0	32	92	0,251	0,833	0,321
	3	1,00	30/32	94					
	1	-0,33	0/32	0					
3	2	1,00	30/32	94	32	63	0,286	0,584	0,217
	3	-0,33	2/32	6					
	1	-0,33	0/32	0					
4	2		0/32	0	32	86	0,335	1,000	0,747
	3	1,00	28/32	88					
	1	-0,33	2/32	6					
5	2		0/32	0	32	76	0,349	0,584	0,022
	3	1,00	30/32	94					

Table 7. Details of test parameters for the study. Test Paired, Topic 10

Nº Questions block	Value of each possible response separately	Nº of times responded /Total Nº responded. For questions	Nº of the questions block	Value of each possible response separately	FI (%)	SD	DI	DC
		38/39	97					
		38/39	97					
	1,00	38/39	97					
		38/39	97					
1		38/39	97	39	94	0,171	0,977	0,861
		1/39	3					
		1/39	3					
	0,00	1/39	3					
		1/39	3					
		1/39	3					
		39/39	100					
		39/39	100					
	1,00	38/39	97					
2		36/39	92	39	90	0,171	0,903	0,771
		36/39	92					
		1/39	3					
	0,00	2/39	5					
		2/39	5					

For test 2, there is a great divergence between the two clearly defined groups of SD. Because of that, it would be correct to change the form in which the questions have been grouped (paired questions), and thus, the four blocks of items would be centered, improving the verification uniformity of the acquired knowledge. The graph shows the area discussed.

Another interesting parameter for the analysis of results of the test, is the Discrimination Index (DI); which provides an approximate indicator of each item (question)

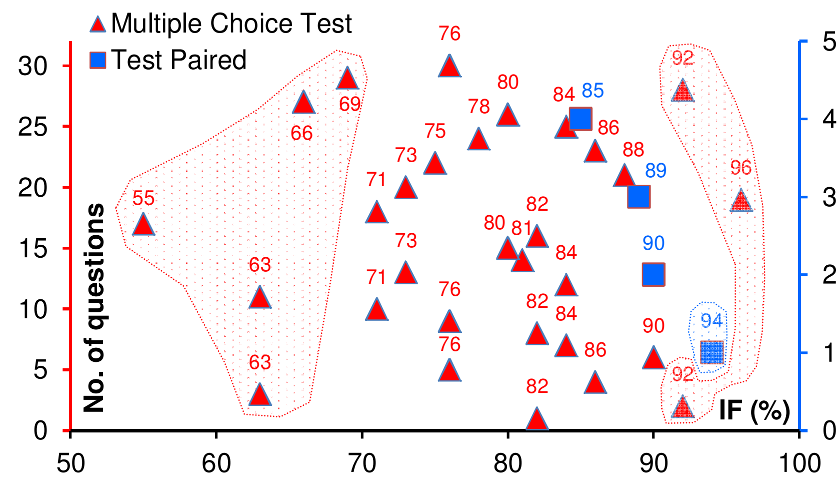

Figure 1. Facility Index for two tests studied

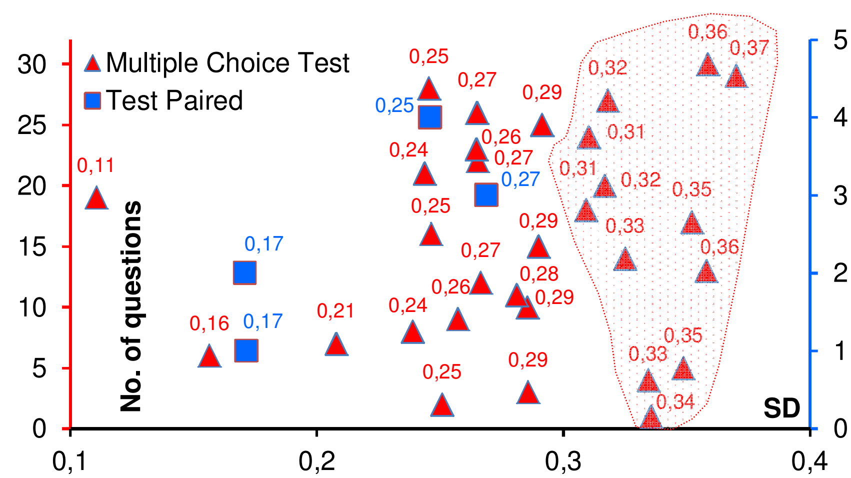

Figure 2. Standard Deviation results for the two test study

or analyzed response (separately) of its performance with regard to the answer with smaller performance level; permitting in this way to deduce between high punctuation with respect to the global punctuation, and user less-expert with respect to the experienced.

This parameter is obtained dividing by thirds the student group analyzed, keeping in mind its scorings with reference to the global questionnaire. Below, for the groups superior and inferior is obtained the average punctuation from the analyzed item (continuing the performance order of up downward); and finally, to the previous value is subtracted the average of the punctuation. The mathematical expression is (2):

Where: $$DI = {(X_{top} - X_{\text{bottom}})}/{N} \qquad (2)$$

X_{top}, is the sum of the reached fraction (obtained/maximum) for this item, for a third of students with higher qualifications in the whole questionnaire; this is, the number of correct answers in this group.

X_{bottom}, is the analogous sum for the students located in the lower third of the questionnaire.

This parameter has values in a range of + 1 to -1, its meaning should be interpreted as: when the Discrimination Index (DI) is getting greater than 0.0, it would mean that more low-performance students have been assumed better this item that students with higher performances. Therefore, these items, as questions for evaluation, should be eliminated for being inadequate; in fact, these items reduce the global score precision of the test.

In our work (Figure 3), and wanting to validate an evaluative questionnaire, about this parameter; in test 1 it

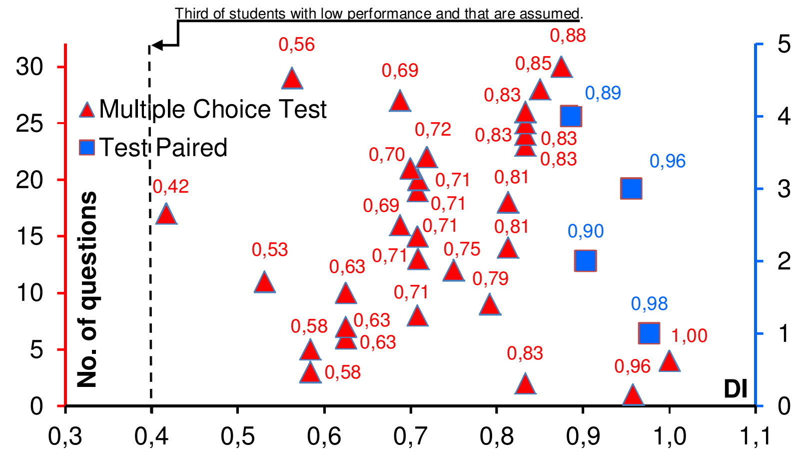

Figure 3. Discrimination Index results for the two test study

will be necessary to eliminate the questions with an DI lower than 0.4; since these are located in the third of students with low performance and that are assumed. It is important to note that, in this case, these questions are not bad designed, but they are not necessary for their evaluation because of their simplicity. The graph shows the border discussed.

In test 2, the concepts before established for test 1 are applicable, completing this questionnaire, with exceeded reliability for future applications. Therefore, it is necessary to adjust the test for application in new practices.

The last statistical parameter analyzed in this work is the Discrimination Coefficient (DC), it is considered as another parameter of measure to achieve the separation of adequate items and low-performance items from the learning evaluation.

The DC is a coefficient of correlation among the scores of each particular item with respect to the complete questionnaire. Its mathematical expression is (3):

$$DC = \sum(xy) \Big/ (N * S_x * S_y) \quad (3)$$

Where:

$\sum_{(xy)}$, is the products summation of the deviations for the samples marks of items, with reference to the total survey or test.

N, is the number of answers obtained for a question or item.

S_x, is the standard deviation value of the results for the fraction of the question.

S_y, is the standard deviation value of the results of the total questionnaire.

As in previous parameter (DI), DC can obtain a range of values among + 1 to -1. Positive values indicate items that discriminate right questions, while indices with negative values are items that are answered by low-performing students. This means that items with negatives DC are answered incorrectly by students, which penalizes the majority of students. Therefore, these topics or test questions must be removed.

The advantage of DC with respect to DI is that the first one utilizes all the population of the analysis group to obtain information for its decision, and not just the extreme upper and lower thirds as DI does it. Consequently, this parameter can be considered as more sensitive to detect the items or questions performance. In our case, Figure 4, the detection of the ineligible questions to be considered in future versions of tests is more evident than with the previous parameter DI. The graph shows the area discussed.

For test 1, where besides the question 17 detected by the DI parameter, the questions 5, 6, 7, 11, 19 and possibly 29 also report serious problems in their resolution by part of the students. For test 2, the only difference among in the use of DI with regard to the values reported using DC, is the value reached in its scale, as well as its higher proximity to the nil value. On the other hand, DI and DC describe similar order and relation.

To finish, although the following comments are out of the scope of the statistical analysis of the test in this work, Figure 5 shows our case average time employed in the resolution of the test with reference to the average grade reached by students. In general conditions, and for the case of test 1, there are scores with high values unrelated with the time spent in the resolution. This fact could be employed to detect concepts used in the learning, such as brilliance and efficacy. On the other hand, students with low scores, who recognize their knowledge deficiency, decline to use adequately all the available time to resolve the questionnaire. In the case of test 2, the students that reaches high or medium scores does not utilize the total available time (up to 27 minutes),

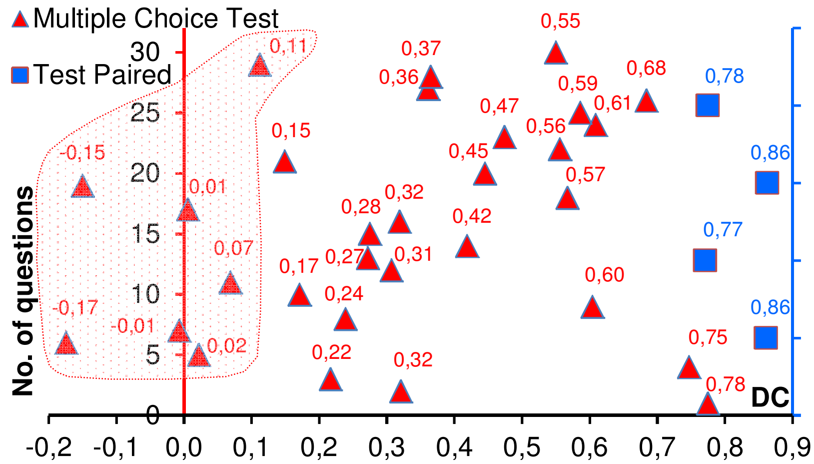

Figure 4. Discrimination Coefficient results for the two test study

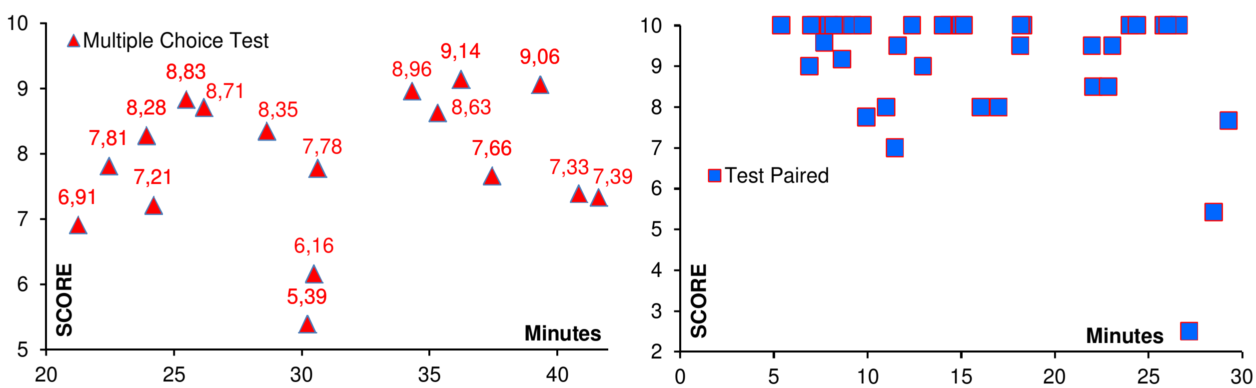

Figure 5. Test resolution study VS. Time resolution

meanwhile students with low scores used it. It is evident that in this test, the resolution time should be adjusted downward, in order to adapting better its use and evaluation.

In order to verify the adaptations, modifications and replacement proposed at both tests, we performed a second evaluation to students of the following year. For this occasion, the thematic content and teacher were the same, but students do not. The comparative analysis of pre and post-test, we observed improvement in the control parameters:

Pre-test:

1. Test 1 FI between 70% to 90% (average 74.47%); Test 2 FI between 85% to 90% (average 89.5%).
2. Test 1 SD average = 0.28; Test 2 SD = average = 0.21
3. Test 1 DI average = 0.73; Test 2 DI average = 0.93
4. Test 1 DC average = 0.33; Test 2 DC average = 0.82

Post-test:

Test 1 FI between 69% to 81% (average 72.1%);

1. Test 2 FI between 75% to 84% (average 73.5%)
2. Test 1 average SD = 0.25; Test 2, average SD = 0.22
3. Test 1 DI average = 0.79; Test 2 DI average = 0.93
4. Test 1 DC average = 0.46; Test 2 DC average = 0.85

It is important to clarify that if this process of analysis of the test could have as many respondents, the level of uncertainty in interpreting the results would be more satisfactory.

CONCLUSION

The final general comments are:

a) At Moodle platform, the tool "Questionnaires", allows faculty a possibility to implement active learning and self-

b) learning experiences with educational purposes. It is also a simple-use instrument suitable for evaluations of the knowledge level reached by students.
c) The use of the available questionnaires at this platform is a big versatility tool, with applications in educational aspects, such as: self-learning, learning evaluation and as criterion of particular adaptation in provided teaching.
d) This tool allows promoting learning activities outside the classroom, reduction in evaluation times (especially in big groups of students) and detection of specific or particular needs of a student or group of students.
e) The implementation of this tool requires extra work of the teacher at the beginning of its use. This initial effort is compensated with the satisfaction in cover the predicted educational expectations, improvements in the educational level reached, and the acceptance of its use on by students.
The specific final comments of this work are:
a) The processed information obtained in tests can contribute with "extra information" which allows adapting all the teaching process in a better form.
b) The Facility Index (FI) permits to discern among the difficulty levels of the questions established in a test; so, it can be used as a criterion to select questions, and thus, to guarantee the adaptation of each one of them, or in lack of that, a scrupulous review of its logic.
c) The Standard Deviation (SD) permits to detect the knowledge acquisition by students. This parameter has a general and uniform character for all the members of the group (general collective knowledge of the theme). Thus, it contributes with criteria of what is or what is not assumed by students.
d) The Discrimination Index (DI) permits to detect those questions to be eliminated in tests because being inadequate for evaluation, thereby improving the precision of the global score of the test. It is important to remark that these questions are not bad designed, but they are not necessary to evaluate because their simplicity.
e) The Discrimination Coefficient (DC) permits to obtain a parameter with detection of ineligible questions in a tests, this is a more sensitive parameter than DI; its use permits to select with success those items more adequate for the knowledge evaluation of students.
f) The control and analysis of the time employed in the evaluation test can contribute with adjustments and additional information of all the evaluation process.

ACKNOWLEDGEMENTS

The authors thank to: Projects Financing Aids for the of the Projects Aid to the Teaching Assembly 2010 by the Sciences of the Education Institute (ICE) UPC and “Atenea Labs”- ICE.

REFERENCES

Batrinca G, Raicu G (2010). Considerations about effectiveness and limits of computer based training in maritime industry. Proceedings of the 3rd International Conference on Maritime and Naval Science and Engineering. 15-20. ISSN: 1792-4707. Romania.
Berrios G., Rojas C, Cartaya N (2005). Effect of the number of options on the quality of est reading comprehension multiple-choice exams. Paradígma. 26(1), 89-116. ISSN 1011-2251.
Blanco M, Ginovart M (2009). Creating Moodle quizzes for the subjects of mathematics and statistics corresponding to the first years in engineering studies. EDULEARN09 Proceedings CD of the International Conference on Education and New Learning Technologies. IATED. 1984-1993.
Blanco M, Ginovart M (2010). Moodle quizzes for assessing statistical topics in engineering studies. Joint International IGIP-SEFI Annual Conference 2010. 1-4. Trnava, Slovakia.
Blanco M, Estela M, Ginovart M, Saà J (2009). Computer assisted assessment through Moodle quizzes for calculus in an Engineering Undergraduate Course. CIEAEM 61 Quaderni di Ricerca in Didattica Matematica. 19(2), 78-83. Montreal, Canada.
Burden P (2008). ELT teacher views on the appropriateness for teacher development of end of semester student evaluation of teaching in a Japanese context. System Science Direct Elsevier. 36, 478-491.
Dougiamas M.-Moodle. Recovered: [December 2011] http://moodle.org/.
Escuela Politécnica Superior de Edificación de Barcelona (2009). Construciones Arquitectónicas II. Universidad Politécnica de Cataluña. Memoria 2007-2008. Barcelona, Spain. Recovered: [December 2011]. http://www.epseb.upc.edu/files/escola/memoria/memoria07-08/curso.pdf.
Esquivel JM (2000). El diseño de las pruebas para medir logro académico: ¿Referencia a normas o criterios? Ravela P. (ed.). PREAL/GRADE. Recovered: [December 2011] http://www.oei.es/calidad2/grade.PDF
Garfield JB (2003). Assessing statistical reasoning. Statistics Educ. Res. J. 2(1), 22-38.
General Public License GNU (2010). Moodle For Teachers, Trainers And Administrators. Recovered: December 2011. http://moodle.org/.
Gómez-Soberón JM (2009). Adaptación de las asignaturas de construcción del Departamento de Construcciones Arquitectónicas II al nuevo concepto de los créditos europeos de Educación Superior y del campus virtual Atenea. Creative Commons. 69. Barcelona. Recovered: [December 2011]. http://hdl.handle.net/2117/8369.
Gómez-Soberón JM, Gómez-Soberón MC (2007). Moodle como herramienta para la creación de un campus virtual: adaptación al EEES. II Jornadas de Enseñanza del Hormigón Estructural. Asociación Cientifico-Técnica de Hormigón. 173-179. Madrid, Spain. Recovered: [December 2011] http://hdl.handle.net/2117/2849.
Gómez-Soberón JM, Bosch-González M, Marín-i-Gordi O, Gómez-Soberón MC, Gómez-Soberón LA (2010). Statistical analysis and use of questionnaires for evaluation of the knowledge at university. A practical case. Ministerio de Educación Superior y Universidad de la República de Cuba. 1-18. La Habana, Cuba. Recovered: [December 2011]. http://hdl.handle.net/2117/9672.
Gómez-Soberón JM, Gómez-Soberón MC, Gómez-Soberón LA (2009). Active learning and autonomous learning: an educational experience. ALE-2009 the learning Experience. 1-11. Barcelona, Spain. Recovered: Creative Commons-UP Commons [December 2011]. http://hdl.handle.net/2099/7809.
Gómez-Soberón JM, Gómez-Soberón MC (2007). Aplicación de una técnica de mejora docente. Jornadas de Enseñanza del Hormigón Estructural. Asociación Científica del Hormigón Estructural. 77-83. Madrid, Spain. Recovered: [December 2011]. http://hdl.handle.net/2117/2848
Heck A, Van Gastel L (2006). Mathematics on the threshold. Int. J. Math. Educ. Sci. and Technol. 37(8), 925-945.
Hutchison D (2009). Designing your sample efficiently: clustering effects in education surveys. Educ. Res. 51(1), 109-126.
Nulty DD (2008). The adequacy of response rates to online and paper surveys: what can be done? Assessment and Evaluation in Higher Education 33(3), 301-314.

Ravela P (ed.) (2000). Los próximos pasos: ¿hacia dónde y cómo avanzar en la evaluación de aprendizajes en América Latina?. PREAL/GRADE. Recovered: [December 2011]. http://www.oei.es/calidad2/grade.PDF

Revuelta J, Ximénez MC, Olea J (2003). Psychometric and Psychological Effects of Item Selection and Review on Computerized Testing. Educational and Psychological Measurement. 63(5), 791-808.

Romero C, Ventura S, Salcines E (2008). Data mining in course management systems: Moodle case study and tutorial. Computer and Education. 51(1), 368-384.

Shepard LA (2006). La evaluación en el aula. Textos de evaluación. Instituto Nacional para la Evaluación de la Educación (Ed.). Distrito Federal.

Tiana A (1997). Tratamiento y usos de la información en evaluación. Calidad y Equidad en la Educación. Iberoamericano Organización de Estados. Programa Evaluación de la Calidad de la Educación. Recovered: [December 2011]. http://www.oei.es/calidad2/tiana.htm.

Tuparov G, Dureva-Tuparova D (2008). On-line testing implementation in open source e-learning environments. International Scientific Conference Computer Science. 875-880.

Van Niekerk J, Von Solms R (2009). Using Bloom's Taxonomy for Information Security Education. Education and Technology for a Better World: 9th IFIP World Conference on Computers in Education (WCCE). 27-31. Bento Gonçalves, RS, Brazil. Recovered: [December 2011]. http://www.wcce2009.org/proceedings/papers/WISE6_Niekerk.pdf.

Veiga A, Amaral A (2009). Survey on the implementation of the Bologna process in Portugal. Higher Education: The International Journal of Higher Education and Educational Planning. 57(1), 57-69.