METHOD FOR QUALITY EVALUATION OF DIGITAL LEARNING TOOLS

R.T. Mora-García, M.F. Céspedes-López, R. Pérez-Sánchez, I. Álvarez-Canteli

University of Alicante (SPAIN)

Abstract

The students´ overcrowding in classrooms and the offer of digital media courses imply the need of new learning objects in education, as these reusable electronic tools allow objective evaluation in large groups by using few resources. The aim of this research is a proposal for a massive assessment in the quality of digital learning tools used in the students´ learning process, by using statistical methods (cluster analysis). This method facilitates the classification and identification of gaps within the assessment and self-learning instruments from different psychometric indicators. The research corresponds to a study case using a learning virtual platform (moodle) where different digital learning objects were implemented and used by students as tools for learning and assessment. Teachers analysed the results applied to objective evaluation and self- assessment tests, which determined whether they were properly designed learning activities and their discriminatory properties. In conclusion, the use of statistical methods massively detected failures or errors in the design of objective tests, allowing an important improvement in the quality and reuse of these resources.

Keywords: Learning Objects, objective test, moodle, discrimination, cluster analysis.

1 INTRODUCTION

The implementation of graduate studies to the European Higher Education Area (EHEA) has allowed a more focused approach in the training of students, measuring the effort that they spend, encouraging their self-learning and favouring more active participation. Studies of college degree should enable the acquisition of those skills needed for the practice of the profession, which must be done by acquiring the knowledge, skills and abilities listed in the respective curricula.

In the field of engineering, it is common to use learning tools intended for self-study and evaluation of the knowledge acquired by students. The use of learning objects in university education has made possible to verify the adequacy and goodness of these reusable electronic tools, in addition to a high acceptance by students.

That is why the current line of work and research conducted by the authors lies in the development of Learning Objects[1] within a virtual learning platform (namely Moodle), which allows great flexibility for easy access to lifelong learning [1] allows its use to a large number of individuals simultaneously and has enabled a quantum leap in the activities of non-face learning.

From the results obtained in previous studies [2], by conducting a perception survey to students coursing the subject "Construction of Non-Structural Elements" at the University of Alicante (sample size *N*=128 students), we learned that the perception on learning with these digital tools was very good, because in all responses a 60% was exceeded with the options Much and Very much (see Fig. 1).

The aspect best valued corresponded to whether the activities had helped them in the learning process of the course, with only 3% of responses indicating Little or Nothing. Feedback on activities, learn new approaches and clear, concise concept activities were also rated positively (93, 92 and 95% respectively). The worst rated aspects corresponded to interest on exercises (if they would do more) with a 13% negative responses (Nothing or Little), and whether the activities were motivating with 12% responding Little or Nothing.

This wide acceptance by students has motivated us to investigate further and to improve learning and assessment instruments used.

[1] There are different names as units of learning, teaching objects, learning objects, instructional objects, learning modules, among others. The term "learning object" is adopted as the most used on the web and major databases [3].

ISBN: 978-84-617-0557-3

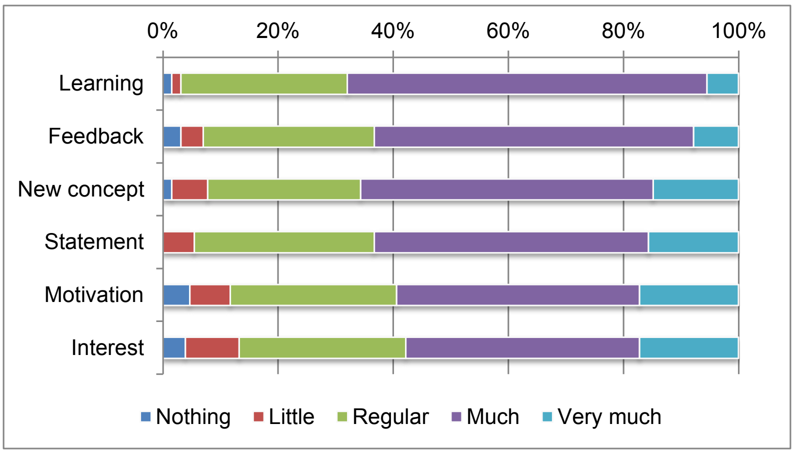

Fig. 1. Perceptions of students on general aspects of the learning objects [2].

1.1 Statement of the Problem

Currently the college classrooms are very crowded, which presents serious difficulties of individual attention to the student. In order to reach the largest possible number of students, the possibility of offering university education digitally opens up. This raises the need for learning objects to evaluate objectively and with few resources a large number of students.

Once learning objects have been developed (hereafter LO) it is necessary to check their validity, psychometric tools exist for it (difficulty level and discrimination power) for determining their goodness thereof.

A case study applied to the subject of Building Non-Structural Elements (hereinafter BNSE) which was conducted in 2011-12 in Technical Architecture and Building Non-Structural Elements course I (hereinafter BNSE I) which was held in the degree in Building Engineering (year 2012-13).

1.2 Objectives

The aim of this research focuses on a proposal to evaluate massively the quality of the instruments used in student's learning by statistical methods (cluster analysis). This method facilitates the classification and identification of gaps in assessment tools and self-study from psychometric indicators. The analysis and treatment of these psychometric indicators have improved the quality of the learning objects used by students.

1.3 Learning Objects

The Institute of Electrical and Electronics Engineers [4] defines "learning object" *as "a learning object is defined as any entity, digital or non-digital, that may be used for learning, education or training*".

There is a lack of consensus on the definition of Learning Object [5] it is also noted that investigations often deal mainly on technological aspects such as accessibility, adaptability, effective use of metadata, reuse and standardization. These authors define LO as "*interactive web-based tools that support the learning of specific concepts by enhancing, amplifying, and guiding the cognitive processes of learners*" [5].

Wiley [6] describes that the Learning Object designers can build small training components (or blocks) that can be reused several times in different contexts, being the digital entity used directly from the internet and by as many users at the same time as necessary.

1.4 Review of Literature

In order to develop appropriate LO it is necessary to know the technical indicators that define the quality of the educational tool used. Typically used, as indicators are the level of difficulty, the discrimination power and the work of the distractors[2]. These are not fixed characteristics of the items, but depend on the group to which the test is applied and will vary depending on aspects such as the

[2] Excluded from this investigation is the functioning of distractors due to its extent.

group ability, the teaching method or even the educational environment. There is extensive literature on the use of psychometric indicators applied to objective assessment instruments. The most important properties of the proposed psychometric attributes are as follows [7]:

1. The item difficulty should not be so high that almost all students cannot answer it correctly, or so easy that most guess its answer, that is, it must have an average degree of difficulty.
2. Items must be able to discriminate between students with higher scores on the exam against lower ranked.
3. Distractors used in the answer options should be appropriate to their purpose.

Description and calculation formula for each of these psychometric indicators is developed below.

a) Facility Index (*FI*): Is the proportion of people who answer correctly an item of a test (% of items answered correctly). This is an inverse relationship: the higher the item difficulty, the lower its index [8].

$$FI_i = \frac{A_i}{N_i}$$

where: FI_i = Item Facility Index i.
A_i = Number of correct answers in item i.
N_i = Number of correct answers plus number of errors in item i.

Table 1. Interpretation of Facility Index.

Facility Index	Interpretation
$FI \leq 0.1$	Very difficult
$0.1 < FI \leq 0.4$	Difficult
$0.4 < FI \leq 0.6$	Medium difficulty
$0.6 < FI < 0.9$	Easy
$FI \geq 0.9$	Very easy

b) Standard Deviation (*SD*): This parameter measures the spread of answers in the response population. If all students respond the same, then *SD* = 0. The *SD* is calculated as the standard deviation for the sample of fractional scores for each particular question.

c) Discrimination index (*DI*): This indicator is intended to distinguish between students of high and low academic performance. It is defined as the degree of differentiation that an item is able to capture through the right or wrong way students respond. Thus, students with higher scores on the test will be those who answer correctly and vice versa.

$$DI_i = \frac{NP_{top} - NP_{bottom}}{N_i/2}$$

where: DI_i = Item discrimination index *i*.
NP_{top} = Number of people with high scores (top third) who answered the item correctly.
NP_{bottom} = Number of people with lower scores (lower third) who answered the item correctly.
N_i = Total people who answered the item *i*.

d) Discrimination Coefficient (*DC*): Is another measure of the power of the item to distinguish the efficient students from the less efficient. The *DC* is a correlation coefficient between scores at the item and the overall score on the questionnaire. The advantage of Discrimination Coefficient against Discrimination Index is that the former uses information from the total student population, not just the extreme thirds (top and bottom). Therefore, this parameter may be more sensitive to detect item performance.

$$DC_i = \frac{\sum(xy)}{N \cdot S_x \cdot S_y}$$

where: *DC* = Discrimination coefficient.

$\sum(xy)$ = Sum of the products of deviations for item scores and overall quiz scores.
N = Number of responses given to this question.
S_x = Standard deviation of fractional scores for this question.
S_y = Standard deviation of scores at the quiz as whole.

Table 2. Interpretation Discrimination Index and Discrimination Coefficient[3].

Discrimination I or C	Discrimination quality	Recommendation
DI or $DC \leq 0$	Very Bad	Discard
$0 < DI$ or $DC \leq 0.2$	Poor	Discard or full revision
$0.2 < DI$ or $DC \leq 0.3$	Regular	Needs revision
$0.3 < DI$ or $DC \leq 0.4$	Good	Improvement possibility
$0.4 < DI$ or $DC \leq 1$	Excellent	Save

2 METHODOLOGY

The methodology used is descriptive and quasi-experimental, by collecting data from the objective tests performed through the Moodle platform during the 2011-12 and 2012-13 courses.

2.1 Description of the context and participants

The population under study is for students enrolled in the subject of BNSE 3rd degree course in Technical Architecture at the University of Alicante during the academic course 2011-12 (N=244) and students of the subject BNSE I degree in Building Engineering Course 2012-13 (N=90).

2.2 Materials

It was necessary to set up the Moodle platform for the use of LO, by which learning activities and questionnaires to collect research data have been developed.

Learning activities were developed as "autonomous modules" [4], as rated by Busetti et al. [9], and have been based on the use of activities with multiple choice questions for self-study and evaluation exercises, and other activities aimed at self-study as crossword puzzles, matching activities and problems combining textual, numeric and multiple choice answers.

Table 3. CENE Activities in the subject.

Activity type	Number of Activities	Number of elements (questions)	Aim
Multiple choice	6	971	Self-study and evaluation
Crosswords	3	47	Self-study
Matching	2	33	Self-study
Problems	3	21	Self-study

The tools used for the preparation of the activities have been Hot Potatoes 6.3 and the Moodle platform. The data collection was carried out from the platform Moodle itself and SPSS was used to analyse them.

[3] The discrimination index and the discrimination coefficient have the same interpretation, and can take values between -1 and 1. Positive values indicate a good discrimination. Negative values are given when items are best returned by students with lower grades o are incorrectly returned by the best students.

[4] As for the different types of Learning Objects the criteria adopted was the one outlined by Busetti et al. [9], which provides some structured LO and others functional. The LO structured may take different forms depending on the approach that the teacher wants to give to the learning process, that is, some "guided modules" (mainly with the active participation of teachers) as "autonomous modules" (where the student must solve a problem or case study, which requires prior basic training) or "hybrid modules" (formed by combination of the above). Functional LO may be of different types depending on the function of its content: "context dependent modules" (containing particular material of the module, eg presentations or assessment modules) and "general purpose modules" (with contents that will work for any module course, eg glossaries).

2.3 Instruments

We used the reports provided by the Moodle platform to calculate the variables under study. First the number of students who answered each multiple choice item, the number of them who have answered correctly, the percentage of correct answers, the Discrimination Index and Discrimination Coefficient for all multiple choice questions.

2.4 Procedures

It was designed a set of activities aimed at self-study and evaluation of students, who could perform them from the Moodle platform in the academic years 2011-12 and 2012-13. With the results of the first year 2011-12 first multiple choice items corrections were made. Some errors were detected and corrected. Once corrected the deficiencies in the items, we proceeded to apply the self-learning tool in the year 2012-13 again, the results of which are analysed in this article.

The results of the multiple-choice items were filtered, refined and codified from Excel. Later, the data was imported into SPSS for descriptive and inferential study. We proceeded first to make a study of bivariate correlation between variables, for later propose an alternative classification that could group homogeneous items on several heterogeneous groups together (cluster analysis).

3 RESULTS

3.1 Descriptive study of variables

Next the descriptive statistics of the variables under study and their coding are described:

Table 4. Descriptive Statistics.

Variable	Coding	N	Min.	Max.	Average	Standard deviation	Skewness	Kurtosis
Facility Index	*FI*	971	0.030	1.000	0.629	0.215	-0.545	-0.472
Standard Deviation	*SD*	971	0.164	0.707	0.437	0.080	-1.174	0.873
Discrimination Index	*DI*	971	-0.667	1.000	0.713	0.237	-1.051	1.150
Discrimination Coefficient	*DC*	971	-0.667	0.839	0.316	0.219	-0.743	1.127

Difficulty Index can take values between 0 and 1, in the case of the Index of Discrimination and Discrimination Coefficient values can range between -1 and 1.

None of the variables meet the normal distribution as seen by the statistics for skewness and kurtosis, having been tested by the normality test of Kolmogorov-Smirnov.

If we look at how the values of the variables are grouped, it can be seen in Fig. 2 that the predominant Facility Index in the items is an easy or very easy (59%) level, compared with a 17% of difficult items and a 24% of medium difficulty. Regarding Discrimination Index the results are very good, with 93% of the items that have good or excellent discrimination (0.3 <DI ≤ 1).

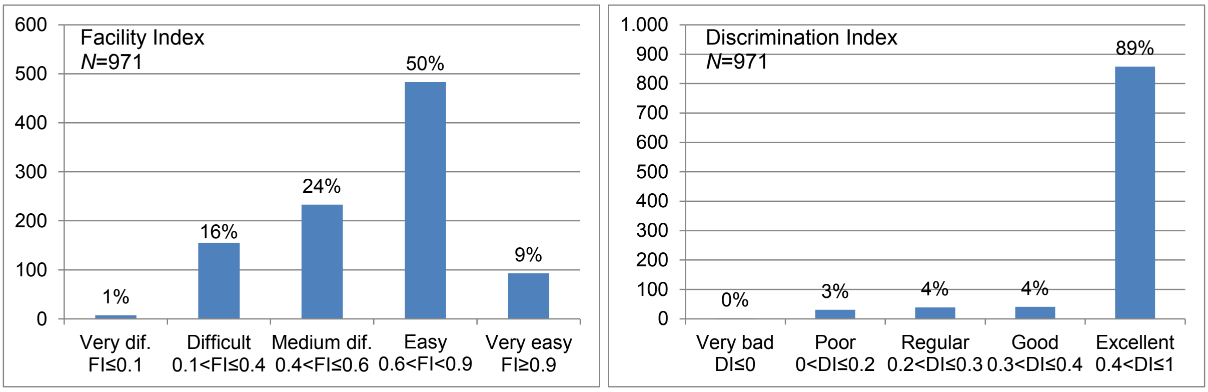

Fig. 2. Diagrams bars of Facility Index and Discrimination Index.

From the analysis of the bivariate correlations between variables, we find that all of them are statistically significant, highlighting the high positive correlation (r = 0.842, p<0.01) between the Index of Discrimination (*DI*) and the Facility Index (*FI*). Moderate positive correlation (r = 0.505, p<0.01;) between Discrimination Index (*DI*) and the Discrimination Coefficient (*DC*) is also highlighted. The correlation between *FI* and *SD*, despite the coefficient *r*, has been graphically analysed and corresponds to a parabolic correlation instead of linear.

Table 5. Matrix correlation Coefficient (Pearson's r), *N*=971.

	FI	*SD*	*DI*	*DC*
Facility Index (*FI*)	1			
Standard Deviation (*SD*)	-0.497**	1		
Discrimination Index (*DI*)	0.842**	-0.309**	1	
Discrimination Coefficient (*DC*)	0.143**	0.085**	0.505**	1

** *The correlation is significant at level 0.01 (bilateral).*

The previous correlations are displayed graphically using a scatterplot in Fig. 3 and 4, the positive trends can be seen in both graphs. The recommendation to identify good discriminant items using the Discrimination Index is to establish a border from *DI* ≥ 0.3 (see Table 2), so lower values should be analysed and studied. Regarding Discrimination Coefficient it can be seen that it is more sensible having taken into account all students. Therefore, we should be more lenient in accepting values below 0.3 *DC*, provided they are greater than zero.

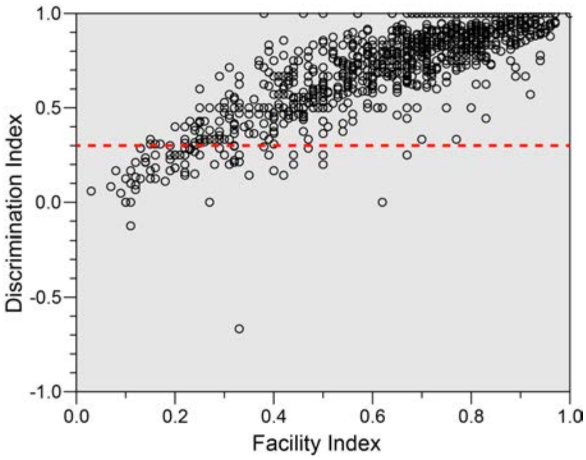

Fig. 3. Scatterplots between *FI* and *DI*.

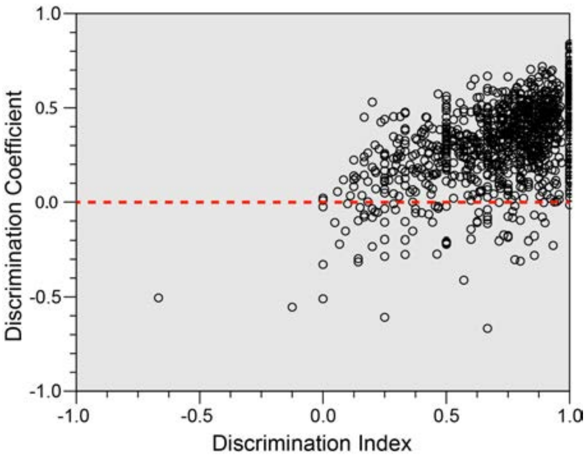

Fig. 4. Scatterplots between *DI* and *DC*.

3.2 Analysis of clusters (cluster)

With the aim to classify multiple-choice items into homogeneous groups, we proceed to perform cluster analysis. The intention is to find a cluster with very similar items within a group (cluster), and secondly that the groups are as most different from each other as possible. This will allow us to identify those items with similar characteristics and analyse them by groups and not individually each item.

The variables to be used are the *FI*, *SD*, *DC and DI,* by a process of hierarchical clustering using the Ward method and the squared Euclidean distance, after standardization of the variables via Z scores.

From the dendrogram analysis of the 971 items, a possible clustering is observed in 4 clusters. Reformulating the analysis, items membership is stored to its respective cluster, allowing for a subsequent classification according to the membership.

The analysis of the scatterplot between variables *DI* and *FI* (see Fig. 5) shows that the clusters are grouped in a nearly ordered manner between them.

Clusters 1 and 2 are those that are located in the intermediate region of the point cloud, and are characterized by a Discrimination Index (*DI*) between 0.3 and 1, as well as a Facility Index (*FI*) between 0.2 and 0.8.

Cluster 3 is located in the upper right end of the diagram and is identified by a high index of discrimination (*M*=0.95) while the items are called very easy (*FI* ≥ 0.9).

Cluster 4 is characterized by its low Discrimination Index (*M*=0.20) and low facility index (most difficult items). These items should undergo a major revision or be discarded.

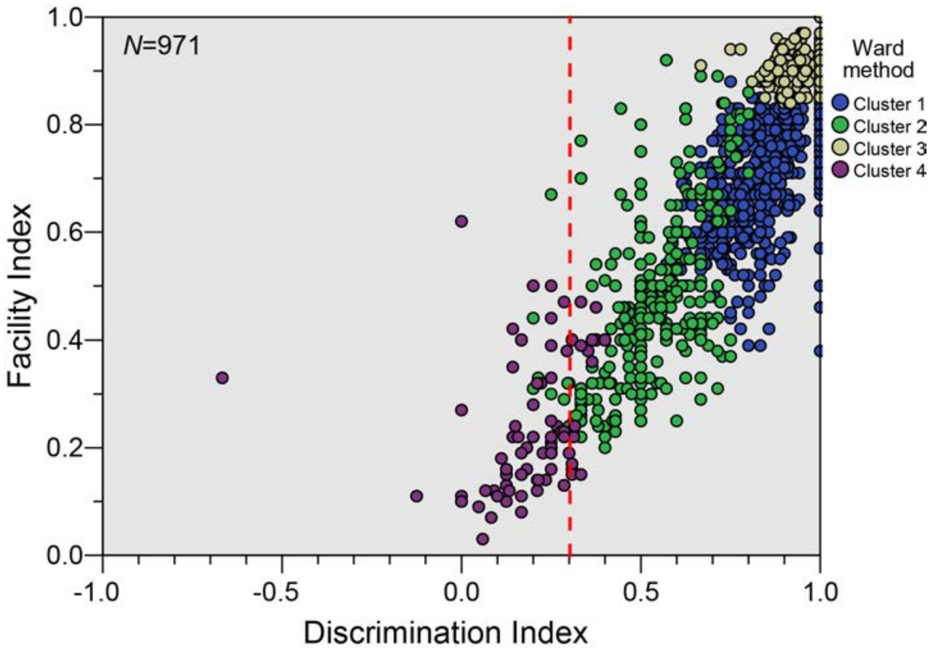

Fig. 5. Scatterplot between *DI* and *FI* by clusters.

When studying the correlation between *DI* and *DC* (Fig. 6) we observe a very similar classification to Fig. 5 but slightly offset in the vertical axis. All clusters behave similarly, lowering the Discrimination coefficient with regards to the Discrimination Index. This indicates the high sensitivity that the *DC* has against the *DI*.

The most contentious items remain those belonging to cluster 4 having *DI* values <0.3 with low or negative *DC*. Therefore, we re-emphasize the need to subject them to a major revision or discard them altogether.

After analysis of these clusters it is necessary to find the reasons why these items are so poor. The main causes identified correspond to items that were poorly assigned the correct answer as well as items on which it has been necessary to make any changes or redrafting.

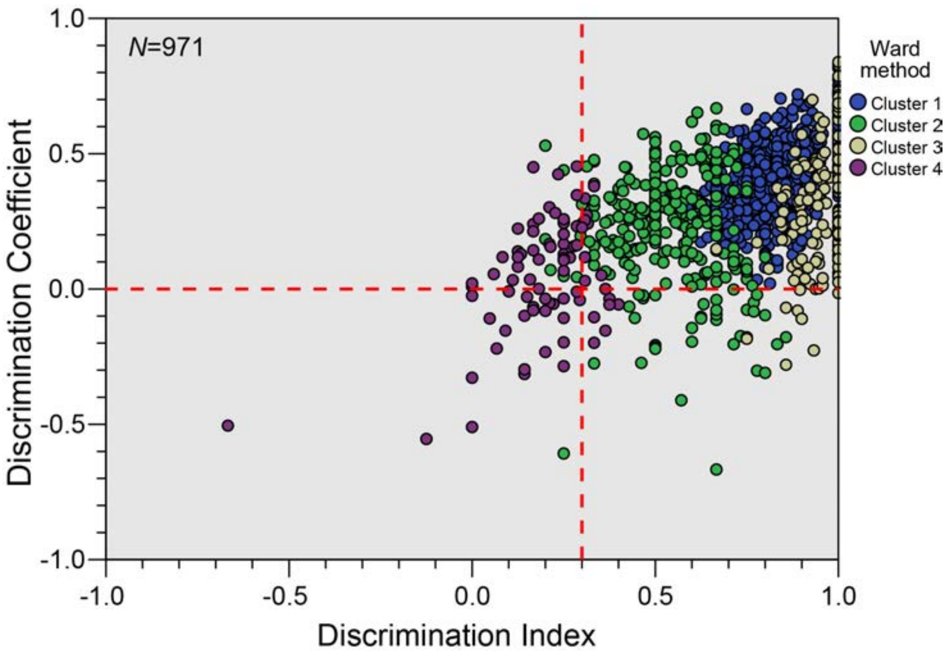

Fig 6. Scatterplot between *DI* and *DC* by clusters.

Table 6. Descriptive statistics cluster.

Variable	Cluster	N	Mean	Standard Deviation	Confidence interval for the mean 95%		Minimum	Maximum
					Lower limit	Higher limit		
Facility Index	1	472	0.694	0.102	0.685	0.704	0.38	0.88
	2	277	0.485	0.162	0.466	0.504	0.20	0.92
	3	142	0.910	0.034	0.905	0.916	0.84	1.00
	4	80	0.237	0.126	0.209	0.265	0.03	0.62
	Total	971	0.629	0.215	0.615	0.642	0.03	1.00
Standard Deviation	1	472	0.458	0.043	0.455	0.462	0.346	0.535
	2	277	0.486	0.040	0.481	0.490	0.289	0.707
	3	142	0.287	0.051	0.279	0.296	0.164	0.374
	4	80	0.409	0.075	0.393	0.426	0.164	0.535
	Total	971	0.437	0.080	0.432	0.442	0.164	0.707
Discrimination Index	1	472	0.829	0.099	0.820	0.838	0.571	1.000
	2	277	0.546	0.130	0.531	0.561	0.200	0.857
	3	142	0.946	0.061	0.936	0.956	0.667	1.000
	4	80	0.197	0.143	0.165	0.228	-0.667	0.400
	Total	971	0.713	0.237	0.698	0.728	-0.667	1.000
Discrimination Coefficient	1	472	0.416	0.145	0.403	0.429	0.020	0.820
	2	277	0.222	0.216	0.197	0.248	-0.667	0.668
	3	142	0.320	0.224	0.283	0.357	-0.280	0.839
	4	80	0.045	0.212	-0.002	0.092	-0.554	0.453
	Total	971	0.316	0.219	0.302	0.330	-0.667	0.839

4 CONCLUSIONS

The teaching-learning tool that was implemented was good quality but required a deep analysis for improvement. Psychometric tools have revealed multiple response items with errors in their writing or in the choosing of the correct answer and have corrected other irregularities that could lead the students to confusion.

We propose a method that facilitates identification of gaps in assessment and self-learning instruments, and by which all multiple-choice items are grouped based on their psychometric indicators. This technique is ideal in situations where a large amount of information is analysed as the case of this research (971 multiple choice items).

The study of the items by cluster analysis has identified groups of items with similar characteristics intra group while the groups are different. We have identified a group of items that suggested a major revision considering the multidimensionality of psychometric indicators and have avoided analysing item by item with no previous criteria.

It has also been possible to make a classification of items according to their difficulty and power of discrimination between good and bad students. All the improvements introduced in these activities have meant higher quality and stability in the learning and assessment tools used in the course

It is pending the review of how the distractors within each item work, which, due to the large volume of questions used, was not analysed in this investigation.

ACKNOWLEDGEMENTS

This research is based on the findings of the “Research in the use of Learning Object for academic teaching”, and was conducted within the context of the call for proposals issued by the "University Teaching Research Networks Project 2012-2013", supported by the Pro-Vice-Chancellor of Strategic Planning and Quality and the Institute of Education Sciences at the University of Alicante.

REFERENCES

[1] Littlejohn, A. (2003). Issues in Reusing Online Resources. *Reusing Online Resources A Sustainable Approach to eLearning*. London: Kogan Page. 107 p. http://www-jime.open.ac.uk/jime/article/view/2003-1-reuse-01

[2] Céspedes-López, Mª F.; Mora-García, R.T.; Rodríguez-Valenzuela, L.; Toledo-Marhuenda, E.; Jiménez-Delgado, A.; Pacheco-Mateo, M. R. (2012). Investigación en la mejora docente mediante Learning Object. In: *Diseño de Acciones de Investigación en Docencia Universitaria*. Alicante: Universidad de Alicante, pp. 3096-3111. http://hdl.handle.net/10045/25861

[3] Calzada-Prado, F. J. (2010). *Repositorios, bibliotecas digitales y CRAI: los objetos de aprendizaje en la educación superior*. Buenos Aires: Alfagrama. 350 p.

[4] IEEE Learning Technology Standards Committee (2002). *IEEE Standard for Learning Object Metadata. IEEE Standard 1484.12.1*. New York: Institute of Electrical and Electronics Engineers. 44 p. http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf

[5] Kay, R. H. y Knaack, L. (2007). Evaluating the learning in learning objects. *Open Learning: The Journal of Open and Distance Learning*, Vol. 22 (nº 1), pp. 5-28. DOI:10.1080/02680510601100135.

[6] Wiley, D. A. (2000). *Learning object design and sequencing theory*. Doctoral thesis. Provo (Utah): Brigham Young University, Department of Instructional Psychology and Technology. 131 p. http://opencontent.org/docs/dissertation.pdf

[7] Carlos-Martínez, E.A.; Vaca-Gastelum, H.H.; Valdés-Cuervo, A.A.; Castro-López,A. (2009). Propiedades psicométricas del examen de admisión del COSNET. *Educatio*, nº8, pp. 23-37. http://www.educacion.ugto.mx/Educatio/PDFs/educatio8/Martinez%20et%20al.pdf

[8] Wood, D.A. (1960). *Test Construction: Development and Interpretation of Achievement Tests*. Columbus,OH: Charles E. Merrill Books, 134 p.

[9] Busetti, E.; Dettori, G.; Forcheri, P. y Ierardi, M. (2005). Devising a Typology of LOs Based on Pedagogical Assumptions. *Advances in Web-Based Learning - ICWL 2005*, Vol. 3583. Springer Berlin, pp. 375-386. DOI: 10.1007/11528043_39.

EDULEARN14

6TH INTERNATIONAL CONFERENCE ON EDUCATION AND NEW LEARNING TECHNOLOGIES

BARCELONA (SPAIN)
7TH - 9TH OF JULY, 2014

CONFERENCE PROCEEDINGS

Published by
IATED Academy
www.iated.org

EDULEARN14 Proceedings
6th International Conference on Education and New Learning Technologies
July 7th-9th, 2014 — Barcelona, Spain

Edited by
L. Gómez Chova, A. López Martínez, I. Candel Torres
IATED Academy

ISBN: 978-84-617-0557-3
ISSN: 2340-1117
Depósito Legal: V-1602-2014

Book cover designed by
J.L. Bernat