

Evaluation of an Examination in Biochemistry and Molecular Biology

Delgadillo-Gutiérrez Héctor Javier¹, Méndez-Ramírez Ignacio² and Saldaña-Balmori Yolanda³

1. Departamento de Sistemas Biológicos, Universidad Autónoma Metropolitana-Xochimilco, Calzada del Hueso 1100, Col. Villa Quietud, Coyoacán, Ciudad de México, 04960, México.

2. Instituto de Investigaciones y Matemáticas Aplicadas y Sistemas, Universidad Nacional Autónoma de México, Av. Universidad 3000, Col. Universidad Nacional Autónoma de México, Coyoacán Ciudad de México, 04510, México.

3. Departamento de Bioquímica, Facultad de Medicina, Universidad Nacional Autónoma de México, Av. Universidad 3000, Col. Universidad Nacional Autónoma de México, Coyoacán Ciudad de México, 04510, México.

Abstract: This work was aimed at assessing one of the examinations applied to the students enrolled in courses of Biochemistry and Molecular Biology at the School of Medicine, UNAM. We analyzed a final examination in this subject. The test consisted of 80 multiple choice questions. The database was exported to Excel® and then to the SPSS16 statistical software for statistical analyses. The following techniques were used: (1) difficulty index (Pi), (2) discrimination index (Di), (3) discrimination coefficient (rpbis), and (4) Cronbach's alpha. Those questions that complied with 3 of the 4 mentioned techniques were considered acceptable; of the 80 questions, only 25 were accepted corresponding to 31%. The topic with the largest number of accepted questions was Water and pH (75%), and the topics without accepted questions were Bioenergetics and Hormones (0%). It is recommended that the faculty members that elaborate multiple choice examinations must know the subject, and should have a formation in didactics and educational methodology.

Key words: Evaluation of the assessment learning in Biochemistry, items evaluation, learning.

1. Introduction

Historically, the educational evaluation process has been a complex task, because many diverse aspects intervene in it; just to mention some, occasionally teaching has been prioritized over learning, and it is rather the latter which really sets the goals of teaching. Another complex aspect is the previous formation of the student that will allow acquiring knowledge on a determined topic, thus previous knowledge must have been acquired to make the new one viable. Finally, among many other participating factors, we must point out that the affinity between the teacher and the student is a determinant factor because many times the success in acquiring the knowledge depends on this affinity.

Corresponding author: Héctor Javier Delgadillo-Gutiérrez, Ph.D., research field: biological sciences, statistical and research education.

A way of testing that teaching is being performed is by assessing the learning obtained by the students; the performed evaluations must provide evidence of the teaching-learning process. Besides, evaluation is a key component within the educational process as it helps to assess whether the proposed objectives have been achieved. Jaap et al. [1] indicated that a valid content and theoretical construct is necessary in an evaluation to identify whether the evaluation instrument is adequate for the purpose, that is, it must be a measure of how well the construct is reflected with that instrument, which is to be supported on a logical structure and be congruent with the knowledge, it must be reliable, appropriate, and viable in terms of times, as well as transparent regarding contents and the way of grading

On the other side, educational evaluation measures of the learning are acquired by the student, but, as a

first instance, this must serve the student to realize whether his/her study technique is efficient or should it be changed by another assumed to provide better learning results. Or rather that the study must be intensified, supported by other materials that will help to comprehend the basic issues, etc. In a third instance, evaluation is aimed at letting the school system know about the student's capacity and from there make decisions on promoting the student.

Those in charge of the evaluation are the ones to decide, those that define what is normal, relevant, adequate or good, in relation to the behavior of students, as well as on the contents that have to be covered, results of their learning, and the times in which the learning must be acquired [2].

The endpoint of an evaluation leads to the emission of a grade, which must be objective, express a correct correspondence with the assimilated knowledge, and diminish the influence of a subjective factor by which the professor might grant different grades to students with the same results.

The evaluation instruments must comply with the requisite that the content is directly related to the educational objectives so that when applied they will demonstrate validity and reliability [3].

There is a great social responsibility in the elaboration of evaluation instruments, as it involves, among others, scholarly success or failure, diversity in the educational yield, good or bad students and teachers, different teaching quality, and the appreciation or not of institutional excellence. The National Center for the Evaluation of Higher Education in Mexico (CENEVAL, for its initials in Spanish) was created with the purpose of regulating the enrollment to the public pre-university schools and universities. It is convenient that all higher education institutions count upon a validated evaluation with high quality standards and should this happen; it would help improve teaching, because those obtained results could indicate which parts of the examination are well structured and where improvements have to

be made to reach excellence, so widely searched for in education.

In 1985, APA (the American psychology association) together with AERA (the American educational research association) and NCME (the national council for the measurement of education) published the manual "Standards for educational and psychological testing" which emphasize how to avoid the misuse of evaluation instruments [4]. This normativity exists in the School of Medicine, UNAM, just like in more developed countries with a high educational standard.

In the School of Medicine of the National Autonomous University of Mexico (UNAM), departmental multiple-choice examinations have been applied in the subjects of the basic cycles simultaneously to all students of one generation since about 50 years. This type of evaluations has advantages when we need to know the knowledge that has been acquired by large students, so the institution can have results right away. Besides they are efficacious to assess whether all groups covered the objectives. However, at the same time they present disadvantages: because of the way in which the questions are written it is not always possible to assess the level of knowledge obtained, besides random correct responses can be obtained, and responses can easily be copied among students, together with the possibility that there might be known questions from previous examinations disseminated among students before the examination.

This paper is presented as part of a project of educational research related to the learning performance by students of Biochemistry and Molecular Biology at the School of Medicine, UNAM. We performed an analysis of the responses given by the students in one departmental examination.

2. Materials and Methods

This work was performed based on the Test of Abilities and Basic Knowledge (EXHCOBA, for its

acronym in Spanish) that is widely used in Mexico to know the quality standards of teaching [5]. The study was retrospective, observational, cross-sectional, and descriptive; hence, it belongs to a descriptive survey [6]. The database was incorporated as follows: those questions that reached the acceptance criterion were marked with 1, and those not were marked with 0. The test was presented by 613 students. Results were exported to an Excel® spreadsheet and then analyzed with the SPSS16 statistical software.

The response variable was the average grade obtained by the students in each of the 80 questions (items), hence, there were 80 study units that were distributed as follows: 6 items of Bioenergetics; 7 of Enzymes, 8 of Proteins, 13 of Carbohydrates, 7 of Oxidative Phosphorylation and Free Radicals, 11 of Lipids metabolism, 20 of Nucleotides and Genetics, 5 of Water and pH, and 3 of Hormones.

Four techniques were used to perform the analysis: (1) Difficulty index (P_i), which is useful to identify the easiness or difficulty of an item according to the given responses; (2) Discrimination index (D_i) in which we analyzed the 27% of students with the highest grade in contrast to the 27% with the lowest grade; (3) Discrimination coefficient (rpbis), which has the advantage of including the dispersion and the average of each item in 100% of students and responses, hence, it is considered the most adequate; (4) Cronbach's *alpha* that consists in evaluating how reliable an examination is in each of its topics and indicates which items have to be eliminated to increase its validity.

3. Objective

To demonstrate that there are ways by which an instrument used to evaluate learning in students that can be graded, by validating the items of the examinations by means of different techniques as Cronbach's *alpha*, Difficult index (P_i), Discrimination index (D_i) and Discrimination Coefficient (rpbis).

4. Results and Analyses

Difficulty index. Difficulty (P_i) is understood as the proportion of students that respond correctly to an item in a test. To perform the calculations, the number of students that answered correctly the item is divided by the total number of students, according to the following Eq. (1):

$$P_i = \frac{A_i}{N_i} \quad (1)$$

Where:

P_i = Difficulty index of item i

A_i = Number of students that responded correctly item i

N_i = Total number of students that responded item i

In this way, if the proportion of students that responded correctly item i is greater than 0.75 it is considered easy, if the proportion of students that responded correctly is lower than 0.35, the item is considered difficult [5]. For the question or item to be considered accepted it must be between 0.35 and 0.75 (Fig. 1); of the 80 evaluated items, 52.5% reached the acceptance criterion.

Discrimination index (D_i). A good item must discriminate between those students that had good grades in the examination and those that obtained poor grades. For this, the following Eq. (2) was used:

$$D_i = \frac{GA - GB}{N} \quad (2)$$

Where:

D_i = Discrimination index

GA = Number of correct responses in item i of the 27% of students with the highest scores in the test.

GB = Number of correct responses in item i of the 27% of students with the lowest scores in the test

N = Number of students in the largest group (GA or GB)

The higher the discrimination index, the better will the items discriminate the students with high and low scores [5]. If all students of GA answer correctly one item and all students of GB answer it incorrectly, then $D_i = 1$, to the contrary, then $D_i = -1$, and if both answer

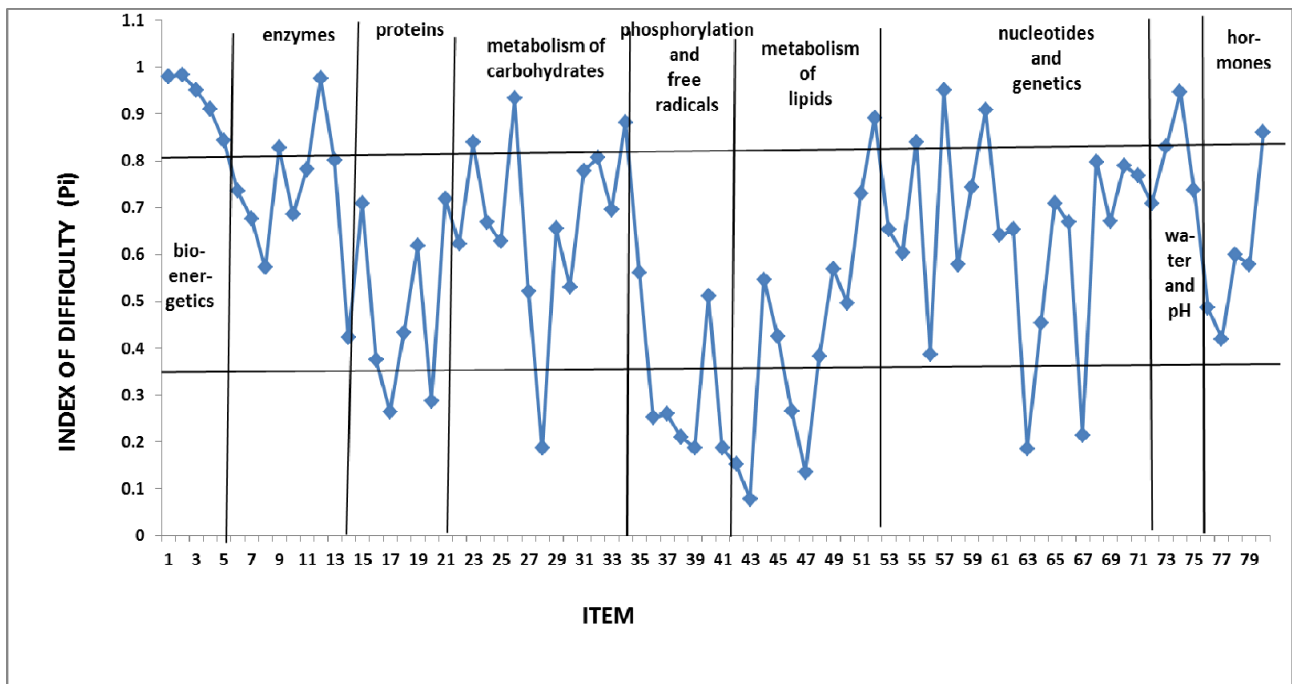


Fig. 1 Difficulty index. The items found in the region between 0.35 and 0.75 were considered accepted.

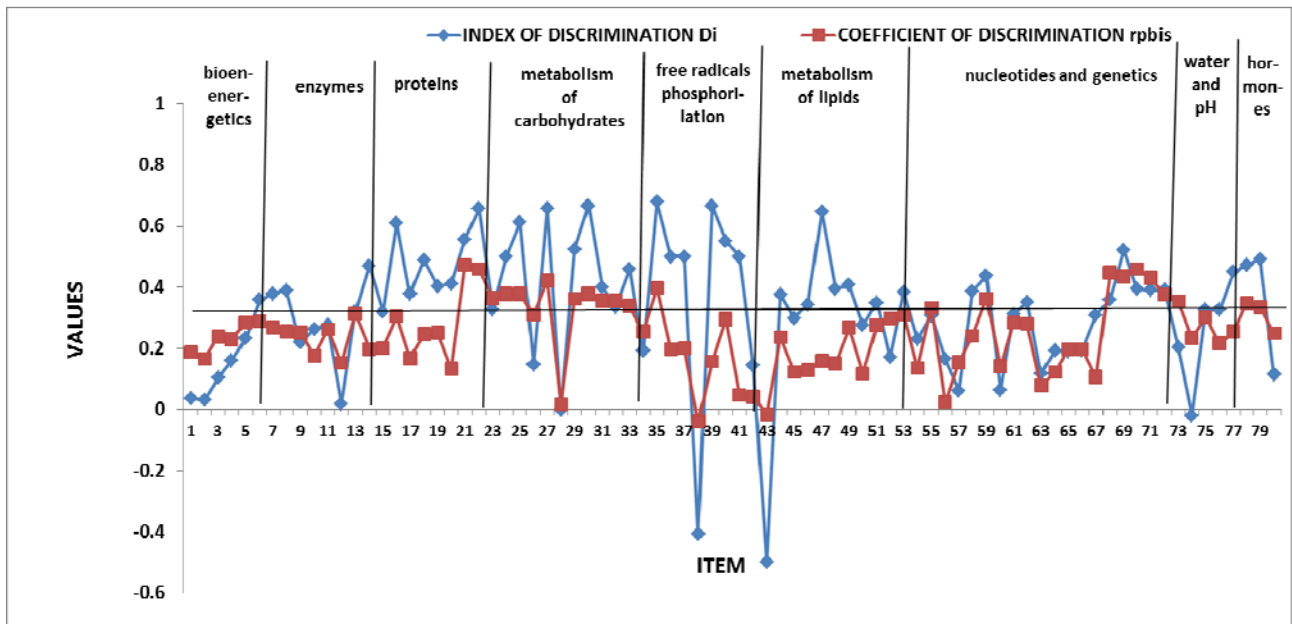


Fig. 2 Discrimination index and coefficient. Items with values above 0.3 in both techniques are considered accepted.

the same, then $D_i = 0$.

For this work, we took as acceptance of the item a value higher than 0.3 (Fig. 2); the following graph was obtained.

Discrimination coefficient (rpbis). In this discrimination coefficient technique, all and each of the evaluated students are taken into account and it is

used to know whether the adequate students are those that obtained the correct responses, what predictive power does the item have, and how it can contribute to the predictions [7]. It suggests that rpbis refers to the predictive validity of the test and is a measure that combines the relation between the criterion of the item and the level of difficulty. Those items that had a

value above 0.3 were considered satisfactory (Fig. 2).

The equation to obtain this indicator is Eq. (3) [7]:

$$rpbis = \frac{\bar{x}_1 - \bar{x}_0}{s_x} \sqrt{\frac{n_1 n_0}{n(n-1)}} \quad (3)$$

Where:

\bar{x}_1 = mean of the total score of those students that responded correctly the item.

\bar{x}_0 = mean of the total score of those students that responded incorrectly the item.

s_x = standard deviation of the total scores.

n_1 = number of cases that responded correctly the item.

n_0 = number of cases that responded incorrectly the item.

$$n = n_1 + n_0$$

Cronbach's *alpha*. Cronbach's *alpha*, also named reliability coefficient, is based on the internal consistency of a test or examination [8]. In this study, it served to see how reliable is the knowledge of a student in the subject of Biochemistry and Molecular Biology. The value of the reliability coefficient varies between 0 and 1, where 1 indicates maximal stability or consistency in scores, this value increases if the

items that had a very low total correlation with respect to the other items of the test are discarded. According to the obtained data, it would be convenient to eliminate items 28, 38, 43, and 56 among others, because if they are eliminated the value of *alpha* would increase, therefore, the exam would increase its reliability (Fig. 3)

Fig. 4 presents an integration of the valid items because they comply with at least 3 of the 4 proposed techniques. Of the 80 items of the test, only 25 complied with the acceptance criterion (31.25%).

5. Discussion

The results indicate that a large percentage of the students subjected to this examination in Biochemistry and Molecular Biology of the School of Medicine, UNAM, did not obtain satisfactory grades. One of the questions that arise systematically is "why is there such a high degree of non-approved students among the freshman population". Actually, this is related to many factors. Despite that the student that enrolls in this School comes from the preparatory school of the UNAM with an average grade of 9.0, where the highest is 10.0, or the student might come from another

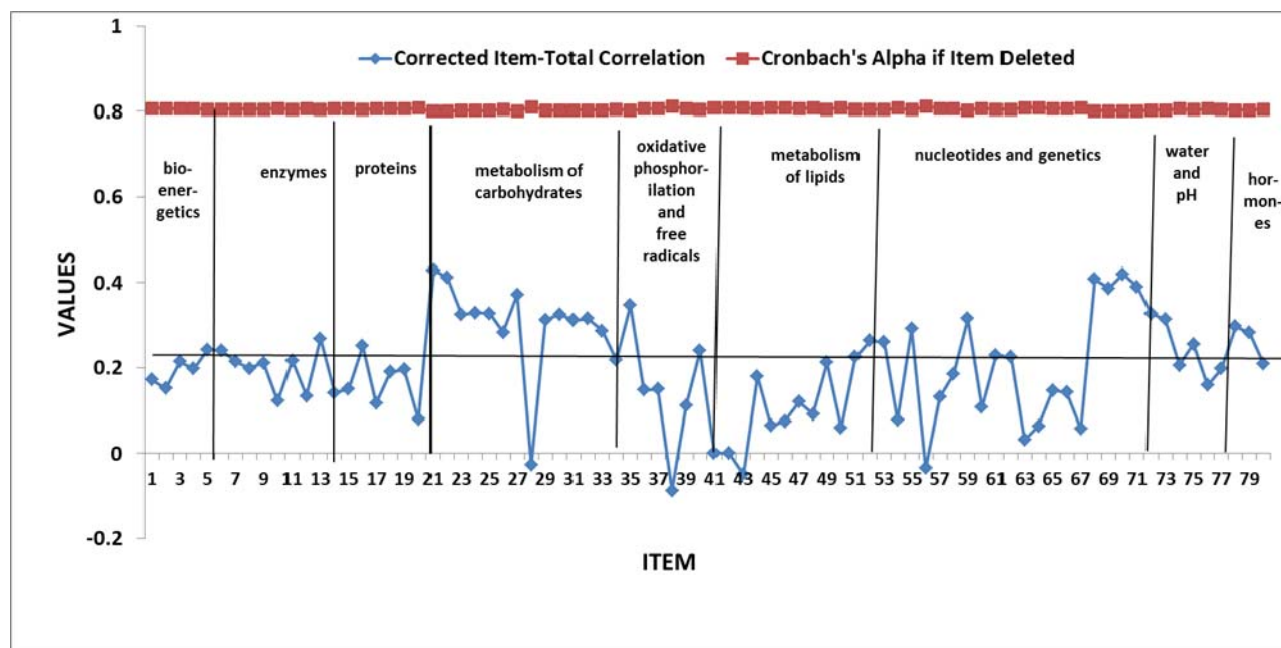


Fig. 3 Cronbach' *alpha*. This value increases if the items that had a very low total correlation with respect to the other items of the test are discarded.

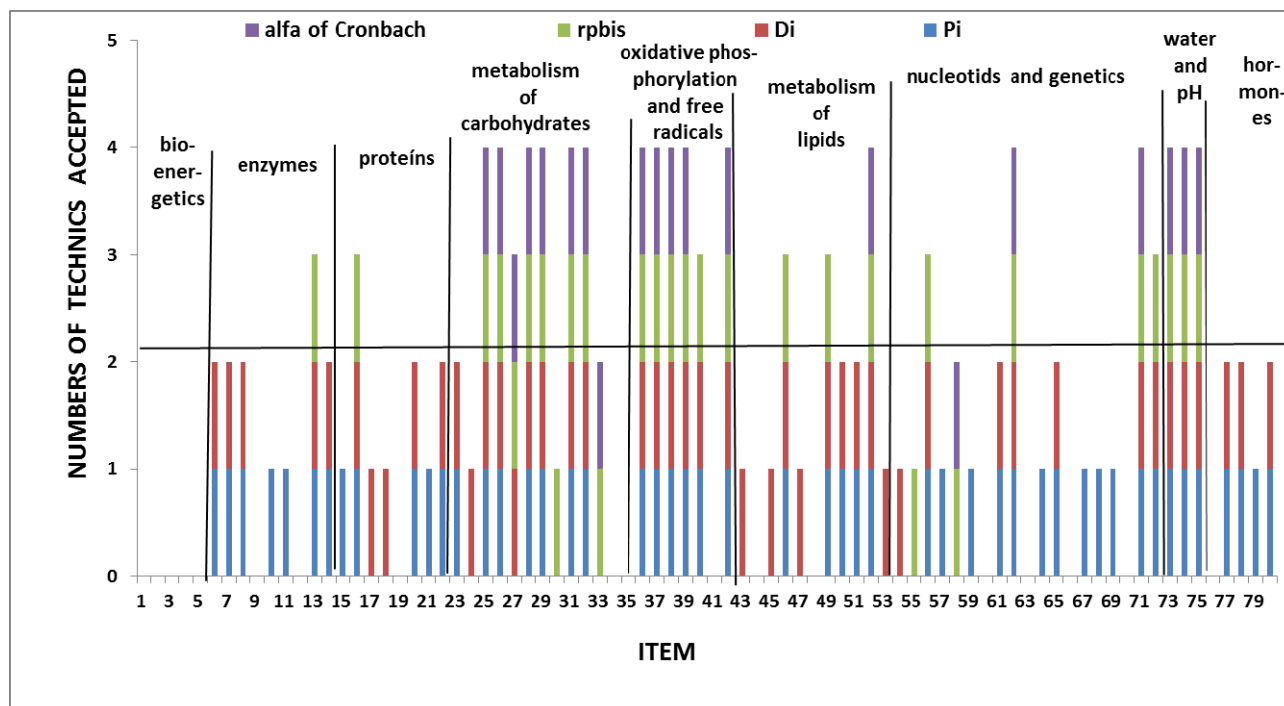


Fig. 4 Techniques used to evaluate the examination. Items with 3 or 4 converging bars are considered as well elaborated and reliable; it can be observed that only 25 of the 80 items comply with the requisite.

type of preparatory school but only after crediting an admission exam, and only a very small proportion of students are accepted (only the top highest grades). The possible responses to this question are related to the change in the school system, level of exigency, complexity of the contents of the subjects of the School of Medicine, different study techniques, and issues related to the latter when going from the preparatory school to the university level, as demands are increased. Another possible factor is that in some of the disciplines of the preparatory school, the students did not acquire the indispensable knowledge for the subjects coursed at the university level, either because they are not included in their programs or due to the possible tolerance of some teachers in granting grades that will help them to enter the School of Medicine.

After observing the results, the most significant concern is: Why there were so many questions that did not reach the quality standard that was asked for? A possible response is that perhaps the faculty members that elaborated the exam did not have a large enough

database with validated items to choose those that assessed best the contents of the program. Another possibility is that the items used were elaborated for the occasion, without taking care of contents, syntax, writing, pertinence, etc.

All of this leads to the need of counting upon a database with validated items to use the most adequate in the different departmental examinations of the different subjects. This will grant more reliability of the true grades obtained by the students of either Biochemistry and Molecular Biology or any other discipline.

6. Conclusions

A finding that the teaching process is carried out correctly is through an appropriate assessment, in which items about most of the topics included in the program, added to the number of items selected in the different themes proportional to the importance and dedication during the course time should be developed. In this work there were topics as bioenergetics in which none of the six items was considered suitable

for evaluation of a student population that applied the evaluation instrument because they had a high rate of difficulty.

It is possible that based on the criteria of acceptance of the questions the four techniques were used herein. A good option might be obtained to adequately choose the questions to be used in the examinations applied to the students of the School of Medicine, UNAM. If in each test a scrutiny is performed, it is possible to attain a large acceptance range of the used items; if a good database of questions were available, then a higher quality would be obtained in the evaluations that are applied at the School of Medicine, UNAM. Ideally the exam questions should be formulated in according with the topics. The number of questions in each evaluation must be proportional to the importance of each topic. Ideally, it should pass at least three of the four techniques used in this study.

Acknowledgment

Thanks to the Secretary of Scholar Services of the School of Medicine, UNAM for providing the database containing exam results.

References

- [1] Jaap, S., Cees, G., and Sall, M. T. 2003. *Educational Evaluation, Assessment, and Monitoring: A Systematic Approach*. The Netherlands, Lisse, Swets and Zeitlinger Publishers.
- [2] González, P. M. 2000. "Evaluación del aprendizaje en la enseñanza universitaria." *Revista Pedagógica Universitaria* 5 (2): 31-55.
- [3] Días Rojas, P. A., and Leyva, S. E. 2013. "Metodología para determinar la calidad de los instrumentos de evaluación." *Educación Médica Superior* 27 (2): 269-86.
- [4] Standards for Educational and Psychological Testing. 1985. 160 Standards Organized in Four Sections.
- [5] Backhoff, E., Larrazolo, N., and Rosas, M. 2000. "Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA)." *Revista Electrónica de Investigación Educativa*, 2 (1). Accessed November 25, 2016. <http://redie.uabc.mx/vol2no1/contenido-backhoff.html>.
- [6] Méndez, R. I., Namihira, G. D., Moreno, A. L., and Sosa de, M. C. 1990. *El protocolo de investigación. Lineamientos para su elaboración y análisis*. Segunda edición. México: Editorial Trillas.
- [7] Henrysson, S. 1971. "Gathering, Analysing, and Using Data on Test Items." In *Educational Measurement*, edited by Thorndike R. L. Washington, DC: American Council on Education.
- [8] Camacho, R. J. 2003. *Estadística con SPSS para Windows*. México: Alfaomega Ra-Ma.