

# Design and Application of a Computer Tool to Evaluate the Goodness of Fit for Tests Designed to Be Self-Taught <sup>†</sup>

I. T. Molina-Martínez \*, V. Andrés-Guerrero, I. Bravo-Osuna, R. Ruiz-Caro, P. Pastoriza, M. D. Veiga-Ochoa, R. Herrero-Vanrell and M. E. Gil-Alegre \*

Department of Pharmacy, Pharmaceutical Technology and Food Technology. Faculty of Pharmacy, Complutense University of Madrid, 28040 Madrid, Spain; vandres@ucm.es (V.A.-G.); ibravo@ucm.es (I.B.-O.); rruizcar@ucm.es (R.R.-C.); ppabal@ucm.es (P.P.); mdveiga@ucm.es (M.D.V.-O.); rociohv@ucm.es (R.H.-V.)

\* Correspondence: iremm@ucm.es (I.T.M.-M.); megil@ucm.es (M.E.G.-A.)

<sup>†</sup> Presented at the 2nd Innovative and Creative Education and Teaching International Conference (ICETIC2018), Badajoz, Spain, 20–22 June 2018.

Published: 31 October 2018

**Abstract:** Nowadays, multiple choice questions play an important role in the self-evaluation process. The present work undertakes the design of a computer tool—Excel® worksheet—that will allow the calculation of different parameters that are usually employed for the evaluation of tests: difficulty, discrimination index, consistency, etc. The designed tool is used to evaluate the goodness to fit of multiple-choice tests on a practical particular case: self-teaching by competencies on the academic field of Pharmaceutical Technology of the Pharmacy Degree at Complutense University of Madrid. The easy-access computer tool designed makes it possible to evaluate tests from the empirical evidence, with respect to the fulfilment of the desired psychometric requirements, aiming to be useful for the student self-learning.

**Keywords:** psychometric analysis; index of difficulty; index of discrimination; consistency; Excel® sheet; item of multiple choices; Pharmaceutical Technology; self-learning

---

## 1. Introduction

Nowadays, the multiple choice exams play an important role in the evaluation process [1].

The incorporation of the students on the creation of tests, the difficulty of their conception and the importance of their correct design to act as a useful tool or resource for the teaching-learning process lead to the need of individual evaluation of the items to rationalize their selection [1].

Then, it is logical to consider the necessity of relying on a tool which allows us to evaluate individually the quality of each of the questions belonging to a multiple choice exam. Hence, allows us to carry out an appropriate selection of those suited before they actually become part of the battery of questions that will be used to create future exams [2–5].

In this context, the present work undertakes the design and application of a computer tool that would verify the reliability and quality of test exams. Although there are existing computer tools designed for this matter as those developed by Assessment System Corporation y Brooks [6,7], it is proposed to use an easy tool, and for this reason, an Excel® sheet is designed in order to allow the calculation of the different parameters used in the evaluation of the items.

## 2. Materials and Methods

The methodology of the present work involves the selection of contents belonging to the subject Pharmaceutical Technology, the design and verification of a multiple choice exam. Their quality, reliability and utility on the student self-taught are analyzed using the computer tool designed. The evaluated multiple choice exams are then distributed to groups of students studying Pharmaceutical Technology (Pharmacy degree) at the Complutense University of Madrid [5].

### 2.1. Multiple Choice Exams Design

#### 2.1.1. Defining the Content to Be Studied

The amount of detail in the contents is directly proportional to the facility to draft the items and to improve the exam contents' validity.

#### 2.1.2. Table of Specifications of Educational Objectives (TSEO) according to Bloom's Taxonomy with the Following Three Levels of Knowledge

- Basic: on facts and concepts (knowledge and comprehension).
- Medium: on procedures (application and analysis).
- Superior or metacognitive (synthesis and evaluation).

The TSEO is then completed by specifying the amount of items needed to correctly depict each one of the aspects to be evaluated.

#### 2.1.3. Item Creation

Multiple choice questions (PEM) are addressed, paying special attention to those incorrect or distracting options to ensure their attractiveness to people who do not know the correct answer or simply have a superficial knowledge on the subject raised on the item and in turn to result irrelevant to those who possess a good knowledge on the evaluated subject.

### 2.2. Questions' Psychometric Properties

It involves the analysis of the questions, following the multiple choice exams questions' resolution by the students. By using the evaluated computer tool, the following aspects are determined: Basic descriptive statistics, Item's difficulty rate (ID) and marking difficulty rate (IDc), Item's discrimination rate, Relation between discrimination and difficulty, Test verification: error measurement or standard error measurement (SEM), Test verification: reliability. Internal consistency analysis is carried out (Cronbach's alpha-coefficient).

### 2.3. Analysis Decisions

Decision making is addressed towards the removal of items, answers, possible changes on the questions' formulation, etc. To be able to establish a common evaluation criterion, to be able to take these decisions, a relation between quality indicators was created and measures to be taken established. The discrimination rate is an extremely useful parameter when rationally selecting the questions. Table 1 shows those suggested by Ebel in 1965 [5].

**Table 1.** Criteria for the discrimination rate's values [5].

Discrimination Rate's Value	Construction	Recommendation
Less than 0	Awful discrimination	Discard
Between 0 and 0.19	Poor discrimination	Discard or thoroughly revise
Between 0.20 and 0.29	Mediocre discrimination	Needs revision
Between 0.30 and 0.39	Acceptable discrimination	Opportunity to improve
Equal or above 0.40	Excellent discrimination	Keep

A relation between difficulty and discrimination rates exists: if an item is fully easy ( $p = 1$ ), it cannot be discriminating ( $D = 0$ ). Same happens with one fully difficult. Table 2 shows the relation between both rates.

**Table 2.** Relation between the discriminating capacity of an item and its difficulty.

Items Evaluation	Difficulty Rate	Discrimination Rate (Maximum Value)
Very easy	0.91–1.00	0.36–0
Easy	0.76–0.90	1–0.36
Slightly easy	0.51–0.75	1
Slightly difficult	0.26–0.50	1
Difficult	0.11–0.25	0.36–1
Very difficult	0–0.10	0–0.36

### 3. Results and Discussion

#### 3.1. Multiple Choice Exam Design

The TSEO is shown in Table 3: each item is related to a competence.

**Table 3.** TSEO for the domain “Medicament and Pharmaceutical Technology: Introduction”.

Medicament and Pharmaceutical Technology: Introduction	Knowledge	Understanding	Concept Application	Problem Solving	Total Items
1. Definition of medicines		Item 1	Item 4		2
2. Different types of medicines	Items 2 and 3				2
3. Main objectives of pharmaceutical technology	Item 5				1
Total Items	3	1	1		5

#### 3.2. Psychometric Properties

An Excel sheet is designed placing the items and the students involved (Figure 1). Once the exam has been carried out, cells are filled: “1” for correct answers and “0” for wrong answers. The formula for calculating the parameters used to verify the exam (ID, IDC, ...) is placed at the end.

**Figure 1.** Excel sheet showing the values of the evaluating parameters for an exam.

From the students’ answers and the corrected multiple choice exams, was then carried through. Table 4 shows the results of the items’ analysis obtained for the conceptual domain: “Medicament

and Pharmaceutical Technology: Introduction” ( $n$  = number of students). To verify the exam’s reliability, an analysis of its internal consistency was carried out (Table 5). In all cases significant values were obtained, indicating the reliability of the exams.

**Table 4.** Psychometric properties of the items of the multiple choice exam.

Item	Level of Knowledge	Difficulty Rate	Discrimination Rate
TF-1 ( $n = 40$ )	Understanding	0.33— <i>moderately difficult</i> -	0.73 (1)
TF-2 ( $n = 40$ )	Knowledge	0.90— <i>easy</i> -	0.18 (0.36)
TF-3 ( $n = 81$ )	Knowledge	0.52— <i>moderately easy</i> -	0.55 (1)
TF-4 ( $n = 41$ )	Knowledge application	0.73— <i>moderately easy</i> -	0.46 (1)
TF-5 ( $n = 41$ )	Knowledge	0.12— <i>difficult</i> -	0.18 (0.42)

**Table 5.** Multiple choice exam’s quality.

Exam’s Templates for Exam 1	$\alpha$ of Cronbach	Measurement Error
Exam 1-A	0.564	1.13
Exam 1-B	0.681	0.95
Exam 1-C	0.577	1.13
Exam 1-D	0.435	1.19

### 3.3. Analysis Decisions

The defined criteria application leads to the actions shown in Table 6.

**Table 6.** Decisions from the analysis of the items.

Item	Level of Knowledge	Action
TF-1 ( $n = 40$ )	Understanding	Keep
TF-2 ( $n = 40$ )	Knowledge	Keep
TF-3 ( $n = 81$ )	Knowledge	Keep
TF-4 ( $n = 41$ )	Knowledge application	Keep
TF-5 ( $n = 41$ )	Knowledge	Keep

**Acknowledgments:** The current work belongs to the “Innovation and improvement of teaching quality projects” of the Complutense University of Madrid (PIMCD-UCM).

**Conflicts of Interest:** The authors declare no conflict of interest. The founding sponsor—PIMCD UCM—had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

### References

1. Sanchez-Elez, M.; Pardines, I.; Garcia, P.; Miñana, G.; Roman, S.; Sanchez, M.; Risco, J.L. Enhancing Students’ Learning Process through Self-Generated Tests. *J. Sci. Educ. Technol.* **2014**, *23*, 15–25.
2. Galofré, T.A.; Wright, A.C.N. Índice de calidad para evaluar preguntas de opción múltiple. *Revista de Educación en Ciencias de la Salud* **2010**, *7*, 141–145.
3. Gali, A.; Roiter, H.; De Molle, D.; Swieszkowski, S.; Atamañuk, N.; Ahuad Guerrero, A.; Grancelli, H.; Barero, C. Evaluación de la calidad de las preguntas de selección múltiple utilizadas en los exámenes de Certificación y Recertificación en Cardiología. *Revista Argentina de Cardiología* **2011**, *79*, 15–20.
4. Gómez de Terreros, I. Análisis evaluativo de calidad de la prueba objetiva tipo test (preguntas de elección múltiple). *Revista de Enseñanza Universitaria* **1998**, *13*, 105–111.
5. Doval, E.; Renom, J. Desarrollo y verificación de la calidad de pruebas tipo test. *Curso organizado por UB, IL3 e ICE-UB*, **2010**.

6. Assessment System Corporation. Available online: <http://www.assess.com> (accessed on 30 May 2018).
7. Brooks, G.P. "TAP (Test Analysis Program)", Ohio University. Available online: <https://www.ohio.edu/education/faculty-and-staff/profiles.cfm?profile=brooksg> (accessed on 30 May 2018).



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).