sciendo

Research Article

Analysis of Competitive Learning at University Level in Mexico via Item Response Theory

Semei Coronado[1]*

Salvador Sandoval-Bravo[1]

Pedro Luis Celso-Arellano[1]

Ana Torres-Mata[1]

*[1]Departamento de Métodos Cuantitativos,
Centro Universitario de Ciencias Económico Administrativas,
Universidad de Guadalajara,
Zapopan, Jalisco, México
Corresponding author

Doi: 10.2478/mjss-2018-0130

Abstract

This paper presents a study of the multiple choice test from the eleventh knowledge tournament for Statistics I, in order to determine whether it instills competitive learning in university students. This research uses Item Response Theory (IRT). The results obtained show that only 27 students (13.43% of the total number of participants) have an acceptable level of ability (1.03 to 2.58), while the level of ability of the rest of the students is not satisfactory (-1.68 to 0.76). The participants are not a group of students seeking to test their knowledge of the subject or looking for an academic challenge. Better strategies for motivating students in terms of competitive learning must be found.

Keywords: *Competitive learning, Item response theory, Multiple choice test, Logistic model*

1. Introduction

Learning methods aim to prepare students in the resolution of a specific problem or how to deal with specific academic content (Hierro, Atienza, & Pérez, 2014). Regueras et al. (2009) describe how competition among students motivates them to work harder. For their part, (Cantador & Conde, 2010; Lawrence, 2004)) agree that competition encourages learning and increases motivation.

When preparing for a competition, students require a style of learning, one of which is competitive learning (Johnson & Johnson, 2002; Kim & Sonnenwald, 2002; Owens & Straton, 1980), which functions through the competition among students seeking to outdo each other by obtaining better results than their competitors. In fact, (Carpio Cañada, Mateo Sanguino, Merelo Guervós, & Rivas Santos, 2015; Fasli & Michalakopoulos, 2015; Lawrence, 2004; Verhoeff, 1997) describe how competitions challenge participants to give their best, instilling active learning, motivation and self-esteem and, even, motivating weaker students to participate in the activity.

The University of Guadalajara (*Universidad de Guadalajara*, or UdG), through its Department of Quantitative Methods (*Departamento de Métodos Cuantitativos*, or DMC) at the University Center for Economics and Administrative Sciences (*Centro Universitario de Ciencias Económico Administrativas*, or CUCEA), applies competitive learning through the Statistics I Tournament (*Torneo de Estadística I*, or TE_I), which comprises two stages.

Various studies have analyzed knowledge examinations at a university level using Item Response Theory (IRT) to evaluate course topics, analyzing both the quality of the items and difficulties with learning, as well as identifying badly performing students (Awopeju & Afolabi, 2016;

Balmori, Delgadillo, & Méndez, 2011; DiBattista & Kurzawa, 2011; Ingale, Giri, & Doibale, 2017; Mitra, Nagaraja, Ponnudurai, & Judson, 2009; Rao, Kishan Prasad, Sajitha, Permi, & Shetty, 2016; Romero, Rojas, Domínguez, Pérez, & Sapsin, 2015).

From the perspective of the authors of this study, no studies have been identified which analyze a multiple choice test, specifically a knowledge tournament held at a university level. It is for this reason that this study presents, for the first time, a multiple choice test via IRT, which enables the identification of the type of students who participate in the tournament. Said tournament has been able to encourage the development of students' knowledge and abilities through competitive learning (DMC, 2017).

The following section presents the data and the methodology, while Section 3 presents the results. Finally, the last section sets out the conclusions.

2. Data and methodology

2.1 Data

CUCEA is a university center offering thirteen undergraduate degrees in the area of economics and administration. The DMC has organized TE_1 since 2006, which has continued for eleven years with the objective of promoting the learning of statistics in students, developing their knowledge and abilities through competitive learning (DMC, 2017).

The DMC has a statistics academy, staffed by professors teaching Statistics I, among other courses related to statistics. One of the academy committees is responsible for the design and evaluation of the two-stage TE_I test, the first of which comprises twenty multiple choice items while the second comprises ten open items. The data to which this research had access corresponds to the first stage of the TE_I, applied in the second semester of 2017. Students taking the Statistics I course in the semester in which the tournament is being held can participate in TE_I. The questions contained in TE_I are not presented here, at the request of the DMC for reasons of data protection.

Two hundred and one students attended the test, the equivalent of 8.20% of the total number of students (2,450) permitted to participate in the first stage. In the second stage, the twenty students with the highest scores were selected to advance to the second stage, in which the top three places were selected and rewarded with scholarships and books sponsored by various organizations.

The test analyzed in this research is taken from the first stage, due to the size of the population, and comprises twenty multiple choice items. Each item contains a correct answer and four distractors. The following subject areas are covered by the test: descriptive statistics; probability; and, discrete probability distributions. To learn more about the specific subject matter covered by the test, visit http://metodos.cucea.udg.mx/estadistica.php.

A multiple choice test has advantages and disadvantages. Of note among the former are the fact that it evaluates knowledge of the terminology, methods and procedures applied by students. Moreover, it enables the student to identify the application of facts and principles and justify methods and procedures, and reduces the probability of answering questions correctly at random. However, ambiguity could also appear in the questions, which can cause erroneous interpretations in terms of both the questions and the answers (Best & Kahn, 2006; DiBattista & Kurzawa, 2011; Miller, Linn, & Gronlund, 2009; Zamri Khairani & Shamsuddin, 2016).

2.2 Methodology

Certain measures used in education often feature an underlying variable of interest, which quantifies a non-observable aspect and is known as a latent variable or treatment, an example of which being a student's ability in a statistics test. One of education's main objectives is to determine the value of this latent variable (ability), namely, the ability of the student. In order to analyze said variable, this study is based on Item Response Theory (IRT) (Baker & Kim, 2017).

Various studies refer to the advantages of using IRT. For example, IRT focuses more on the properties of the individual items than on the global properties of the tests, by means of a non-linear model which could contain one, two or three parameters, which enables the determination of the best model to be adjusted to the data. The ability scores are given on a scale of $-\infty$ to ∞ or can be transformed on a certain scale, and also have the property of invariance. Furthermore, the parameters for the items and people involved are independent of the sample (Aiken, 1979, 2003; Finch & French, 2015; Furr & Bacharach, 2013; Hambleton, Swaminathan, & Rogers, 1991; K., Hambleton, & Jones, 1993; Muñiz, 2010; Zamri Khairani & Shamsuddin, 2016).

Given the above, the first stage of the TE_I is analyzed using IRT, by means of the Rasch logistical model (RM) (Rasch, 1980). All IRT models, including the RM, express the relationship between the level of the student's latent trait (statistical ability) and the probability of passing a certain element (correctly answering the item), which could be modeled using a logistic model. These models rest on three founding assumptions: (1) monotonicity – the relationship between the trait (latent variable) and the probability of responding to the item is monotonically incremental; (2) one-dimensionality – solely one unique latent trait is being measured via the group of items; and, (3) local independence – when the latent trait is controlled, there is no correlation between the responses to the items (Finch & French, 2015).

The RM uses a binary one and zero codification, with a correct answer equal to 1 and an incorrect one equal to 0. Considering that the TE_I first stage test was applied with J students and consisted of *I* items, x_{ij} can be defined as the score obtained from the $j-th$ student in the $i-th$ item. The above can be established as a logistical model with three parameters (3PLM) (Lord, 1980):

$$P\left(x_{ij}=1\middle|\theta_j,a_i,b_i,c_i\right)=c_i+(1-c_i)\frac{e^{a_i\left(\theta_j-b_i\right)}}{1+e^{a_i\left(\theta_j-b_i\right)}}, \quad (1)$$

where $P\left(x_{ij}=1\middle|\theta_j,a_i,b_i,c_i\right)$ is the probability for student *j* with a score 1 versus 0 in item *I*, a_i is the slope for the curve of the model, b_i is the difficulty of the item, c_i is the parameter for guessing item *i*, and θ_j is the ability parameter for student *j*. However, the RM can also be modeled with both two parameters and one sole parameter.

With two parameters (2PLM) (Birnbaum, 1968):

$$P\left(x_{ij}=1\middle|\theta_j,a_i,b_i\right)=\frac{e^{a_i\left(\theta_j-b_i\right)}}{1+e^{a_i\left(\theta_j-b_i\right)}}, \quad (2)$$

With one parameter (1PLM) (Rasch, 1980), cited in (Sinharay, 2003; Thissen & Wainer, 2001):

$$P\left(x_{ij}=1\middle|\theta_j,a,b_i\right)=\frac{e^{a\left(\theta_j-b_i\right)}}{1+e^{a\left(\theta_j-b_i\right)}}, \quad (3)$$

where *a*, takes the same value for all of the items.

3. Results

The test comprises twenty questions, each of which has a correct answer and four distractors. Figure 1 shows the behavior presented by the participants for each of the questions.

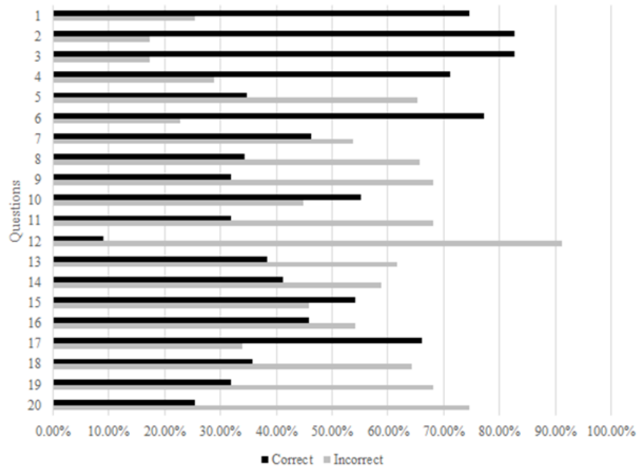

Figure 1. Behavior for correct and incorrect answers per question in the test.
Source: Prepared by the author based on R results.

It can be observed that the first items were answered correctly by a little more than 70% of the participants, after which the percentage increased in the opposite direction, presenting a higher percentage of incorrect answers than correct ones. Therefore, 49% correct answers and 51% incorrect answers were given in the test.

Table 1 presents the descriptive statistics for the test in general, in which a positive asymmetry can be observed, with few students achieving high scores and more students achieving low scores.

Table 1. Descriptive statistics

Mean	9.70
Median	10
Mode	8
Standard deviation	3.05
Sample variance	9.30
Kurtosis	-0.52
Skew	0.31
Minimum	4
Maximum	18

Source: Prepared by the author based on R results.

The scores obtained in the first stage of the test oscillate between 4 and 18 points. The best twenty students are selected to advance to the next stage. In the event that there are more students who have attained the minimum score used to select the twenty best students, all of them pass to the second stage. Therefore, twenty-seven students were selected for the next stage with a maximum score of 18 and a minimum score of 14.

If the scores are placed on a scale of 100 in order to interpret them as a grade, the maximum score was 90 and the minimum was 20, while the minimum for the group that advanced to the next stage was 70. The group that advanced to the next stage corresponded to 13.45%, thus 86.55% obtained a lower grade. The behavior of the grades obtained is presented in Figure 2.

It is possible to speculate that there is only a certain percentage of students who take the test seriously, while the rest appear to have neither the knowledge nor the ability in terms of the subject matter covered by the test, which will be shown by applying the IRT model.

The selection of the best model to be adjusted to the data was carried out using the Bayesiano criteria (BIC), as it provides more parsimonious statistics than those provided by Aikaike (AIC). However, (Finch & French, 2015) describe how these two selection models are not the only ones, identifying distinct approaches for comparing the adjustment of the models, although the best method of adjustment has not been determined.

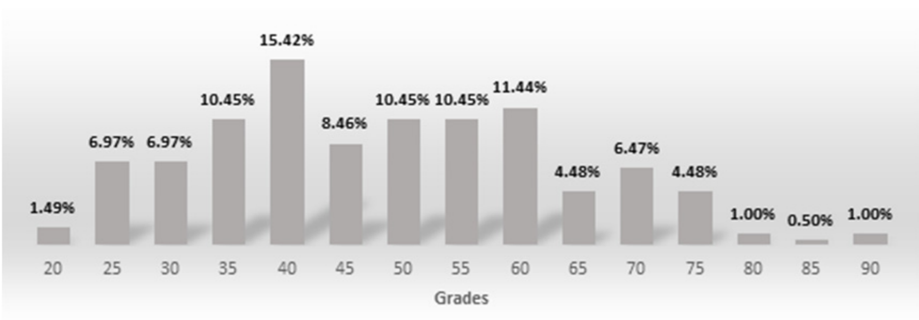

Figure 2. % of students according to their grade
Source: Prepared by the author based on R results.

Table 2 presents the results of the selection of the best model, which is adjusted to the data. It should be noted that, using the 3PLM, it was not possible to perform the calculation due to the fact that the Hessian matrix was not positively defined, for which reason an unstable solution was obtained. Thus, the 1PLM and 2PLM models were compared, with the 1PLM found to be the best model. The calculations were undertaken using the R and Latent Trait Models software package under the IRT proposed by (Rizopoulos, 2017).

Table 2. Statistics for selecting the best adjustment of the model

Model	AIC	BIC
1PL	4793.74	4863.11
2PL	4787.13	4919.26

Source: Prepared by the author based on R results.

The goodness of fit for 1PLM, after 1000 resampling simulations (bootstrap), was $\alpha = 0.05$ and, under the null hypothesis that the model is adjusted adequately to the data with a *p-value*=0.139, the model is, therefore, well adjusted. Thus, each of the items was adjusted to the model. The results are available from the authors if required.

The results for b_i on the test are presented in Table 3 and are ordered from the easiest item (Item 3) to the most difficult (Item 12). These difficulty results coincide with the percentage of questions answered correctly, as shown in Figure 1, from which index both positive and negative values can be taken. The values close to zero express an average difficulty, while the negative values express a difficulty level below the average (low difficulty), and the positive values express a difficulty level above the average (high difficulty). The next column from the table shows the probability of answering the question correctly to the *i-th* item for an average individual, with the probability reducing in accordance with the difficulty of the item (Rizopoulos, 2006).

Table 3. Index of difficulty (b_i) and probability

Item	b_i	Probability	Item	b_i	Probability
3	-2.7859	0.8413	14	0.6333	0.4063
2	-2.7859	0.8413	13	0.8585	0.3743
6	-2.1836	0.7870	18	1.0510	0.3477
1	-1.9424	0.7618	5	1.1293	0.3371
4	-1.6281	0.7260	8	1.1690	0.3319
17	-1.2136	0.6740	9	1.3711	0.3056
10	-0.3827	0.5570	11	1.3712	0.3056
15	-0.3104	0.5463	19	1.3712	0.3056
7	0.2672	0.4601	20	1.9401	0.2384
16	0.3037	0.4547	12	4.1156	0.0784

Source: Prepared by the author based on R results.

The discrimination index must be higher than 35% in order to produce a good discrimination between those students obtaining high and low scores in a test. The index obtained was 59.86% (*coefficient a*) for each of the items, which is acceptable in accordance with (Aiken, 1979, 2003; Romero et al., 2015).

Figure 3 presents the graphs for the characteristic curves for each of the items (ICC), which are ordered from the easiest to the most difficult. For example, the ICC for Item 2 shows that there is a nearly 40% probability of students with a low level of ability in statistics answering correctly. On the other hand, there is an almost 100% probability of students with a high level of ability in statistics answering correctly.

One exception is the ICC for the most difficult item (Item 12), in which there is a low probability of students with a low level of ability answering correctly, while students with a high level of ability have an approximately 50% probability of answering correctly.

The total information test is calculated at an interval of (-10, 10) (Rizopoulos, 2017), while, applying the test at an interval of (0,10) obtained a total for information from the test of 11.87 at an interval of 6.12, which is the equivalent to 51.58%, indicating that 48.42% of the students have an ability lower than 0. This behavior is presented in Figure 4 with the curve for the total information, which is centered at almost zero and is almost symmetrical with a slight bias to the left. From this result, it can be concluded that the test applied in the first stage of TE_I is a test for students of average ability; therefore, it is a tournament in which students with a high level of ability or those interested in an academic challenge do not participate. Thus, it does not comply with the DMC objectives of promoting competitive learning through this activity.

Estimating the treatment variable (ability θ) for the sample using the Kernel density estimation obtains a median close to zero and a positive asymmetry, which resembles the total information curve (see Figure 4).

The ability of the twenty-seven students who passed to the final of the tournament was between 2.58 and 1.03, while the ability of the rest of the participants was estimated at between 0.7586 and -1.6759. Of these remaining participants, 7.96% can be considered to have either guessed the answer or that they have a knowledge deficit (Reise, 1990). Only one student of the twenty-seven that passed to the final was found to have guessed or to have a knowledge deficit.

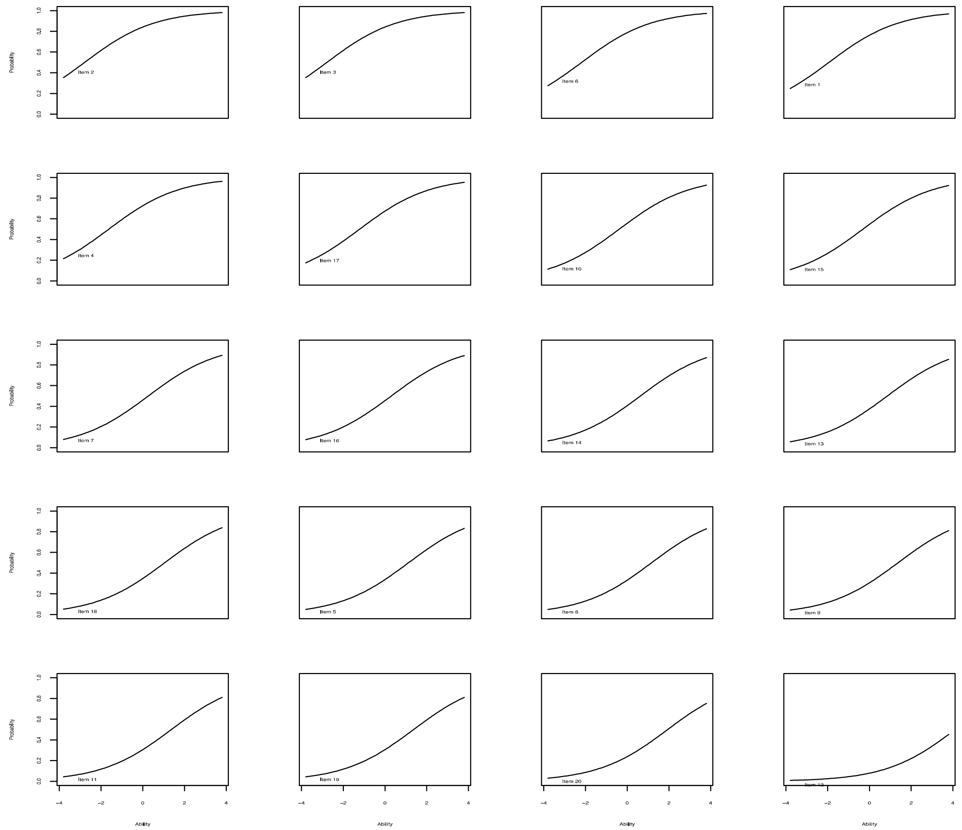

Figure 3. Characteristic Curves for each of the items
Source: Prepared by the author based on R results.

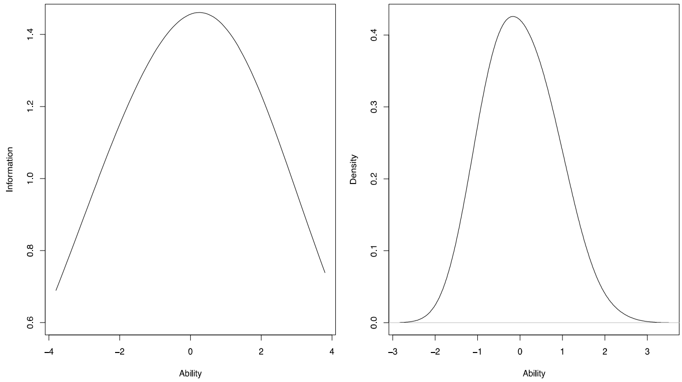

Figure 4. Test Information Function (left) and Kernel Density (right)
Source: Prepared by the author based on R results.

4. Conclusions

Of the studies on partial knowledge tests or those graduating from undergraduate degree programs, various have used IRT (Awopeju & Afolabi, 2016; Balmori et al., 2011; DiBattista & Kurzawa, 2011; Escudero, Reyna, & Morales, 2000; Gajjar, Sharma, Kumar, & Rana, 2014; Ingale et al., 2017; Marie & Edannur, 2015; Mitra et al., 2009; Rao et al., 2016; Romero et al., 2015). However, from the perspective of the authors of this study, this is the first paper to analyze a test taken from a statistics knowledge competition using IRT, by means of which it sought to identify university students' competitive learning.

This research found that the test is designed for students with an average level of ability in statistics, given the low scores obtained and the fact that only 13.43% of students advanced to the next stage of the tournament. However, within this percentage, there was one student who, according to the results obtained, answered the test at random. IRT discriminates between students' distinct levels of ability in a test. Not only should student ability be considered, but also whether the distractors for the test were correctly measured (McDonald, 2017). This aspect could have influenced the results and could be the subject of detailed analysis in future investigations. However, it should be noted that the production of good quality distractors is not an easy task (DiBattista & Kurzawa, 2011).

Through this activity, the DMC seeks to promote competitive learning, with participation in said event low compared to the total number of possible participants. The students who participated are not seeking to test their knowledge of the subject matter nor do they have much interest in it. Better strategies for motivating students to put their abilities to the test must be found.

Furthermore, as the professors who participated in devising the test might not have the ability to design adequate items, they may often produce items that are similar to those found in the textbooks, which may not always be the most adequate for the context. Therefore, it is also recommended that an evaluation of teaching practices is carried out in order that they are consistent with the type of instruments used, which would enable the development of a bank of *adhoc* items for the tournament (Zamri Khairani & Shamsuddin, 2016).

5. Acknowledgments

The authors thank the Department of Quantitative Methods and the Academy of Statistics at the University Center for Economics and Administrative Sciences for providing the data used for this research.

References

Aiken, L. R. (1979). Relationship between the item difficulty and discrimination indexes. *Educational and Psychological Measurement*, *39*, 821–824. https://doi.org/10.1177/001316447903900415

Aiken, L. R. (2003). *Test psicológicos y evaluación* (11th ed.). Naucalpan: Pearson Education.

Awopeju, O. A., & Afolabi, E. R. I. (2016). Comparative analysis of classical test theory and item response theory based item parameter estimates of senior school certificate mathematics examination. *European Scientific Journal*, *12*(28), 263–284. https://doi.org/10.19044/esj.2016.v12n28p263

Baker, F. B., & Kim, S.-H. (2017). *The basics of item response theory using R*. Springer International Publishing. https://doi.org/10.1007/978-3-319-54205-8

Balmori, S. Y., Delgadillo, G. H., & Méndez, R. I. (2011). Evaluación de un examen parcial de bioquímica. *REB. Revista de Educación Bioquímica*, *33*(4), 3–7.

Best, J. W., & Kahn, J. V. (2006). *Research in education*. *Pearson* (10th ed.). Boston, MA: Pearson Education.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. Lord & M. Novick (Eds.), *Statitsical Theories for Mental Test Scores* (p. 592). Reading,MA: Addison-Wesley.

Cantador, I., & Conde, J. (2010). Effects of competition in education: a case study in an elearning environment. In *IADIS international conference e-Learnig* (pp. 11–18). Retrieved from http://arantxa.ii.uam.es/~Cantador/doc/2010/elearning10.pdf

Carpio Cañada, J., Mateo Sanguino, T. J., Merelo Guervós, J. J., & Rivas Santos, V. M. (2015). Open classroom: enhancing student achievement on artificial intelligence through an international online competition. *Journal of Computer Assisted Learning*, *31*(1), 14–31. https://doi.org/10.1111/jcal.12075

DiBattista, D., & Kurzawa, L. (2011). Examination of the quality of multiple-choice items on classroom tests. *Canadian Journal for the Scholarship of Teaching and Learning*, *2*(2), 1–23. https://doi.org/10.5206/cjsotl-rcacea.2011.2.4

DMC. (2017). Torneo de estadística. Retrieved from http://metodos.cucea.udg.mx/estadistica.php

Escudero, E. B., Reyna, N. L., & Morales, M. R. (2000). The level of difficulty and discrimination power of the basic knowledge and skills examination (EXHCOBA). *Revista Electrónica de Investigación Educativa*, *2*(1), 1–16.

Fasli, M., & Michalakopoulos, M. (2015). Learning through game-like simulations. *Innovation in Teaching and Learning in Information and Computer Sciences*, *5*(2), 1–11. https://doi.org/10.11120/ital.2006.05020005

Finch, H., & French, B. (2015). *Latent variable modeling with R*. New York, NY: Routledge.

Furr, M. R., & Bacharach, V. R. (2013). *Psychometrics: an introduction*. *Sage Publications* (Second). Sage Publications, Inc.

Gajjar, S., Sharma, R., Kumar, P., & Rana, M. (2014). Item and test analysis to identify quality multiple choice questions (MCQS) from an assessment of medical students of Ahmedabad, Gujarat. *Indian Journal of Community Medicine*, *39*(1), 17. https://doi.org/10.4103/0970-0218.126347

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory. https://doi.org/10.2307/2075521

Hierro, L. Á., Atienza, P., & Pérez, J. L. (2014). Una experiencia de aprendizaje universitario mediante juegos de torneo en clase. *REDU. Revista de Docencia Universitaria*, *12*(4), 415–436. https://doi.org/10.4995/redu.2014.5634

Ingale, A. S., Giri, P. A., & Doibale, M. K. (2017). Study on item and test analysis of multiple choice questions amongst undergraduate medical students. *International Journal of Community Medicine and Public Health*, *4*(5), 1562–1565. https://doi.org/10.18203/2394-6040.ijcmph20171764

Johnson, D. W., & Johnson, R. T. (2002). Learning together and alone: overview and meta-analysis. *Asia Pacific Journal of Education*, *22*(1), 95–105. https://doi.org/10.1080/0218879020220110

K., R., Hambleton, & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, *12*(3), 38–47. https://doi.org/10.1097/01.mlr.0000245426.10853.30

Kim, S. L., & Sonnenwald, D. H. (2002). Investigating the relationship between learning style preferences and teaching collaboration skills and technology. *Proceedings of the Asis Annual Meeting*, *39*, 64–73. https://doi.org/10.1002/meet.1450390107

Lawrence, R. (2004). Teaching data structures using competitive games. *IEEE Transactions on Education*, *47*(4), 459–466. https://doi.org/10.1109/TE.2004.825053

Lord, F. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Routledge.

Marie, S. M. J. A., & Edannur, S. (2015). Relevance of item analysis in standardizing an achievement test in teaching of physical science. *Journal of Educational Technology*, *12*(3), 30–36.

McDonald, M. (2017). *The nurse educators guide to assessing learning outcomes*. Burlington,MA: Jones & Bartlett Learning.

Miller, M. D., Linn, R., & Gronlund, N. (2009). *Measurement and assessment in teaching* (10th ed.). upper Saddle River, NJ: Pearson Education.

Mitra, N., Nagaraja, H., Ponnudurai, G., & Judson, J. (2009). The levels of difficulty and discrimination indices in type a multiple choice questions of pre-clinical semester 1 multidisciplinary summative tests. *International E-Journal of Science, Medicine & Education*, *3*(1), 2–7.

Muñiz, J. (2010). Las teorías de los test: teoría clásica y teoría de respuesta a los items. *Papeles Del Psicólogo*, *31*(1), 57–66.

Owens, L., & Straton, R. G. (1980). The development of a cooperative, competitive, and individualised learning preference scale for students. *British Journal of Educational Psychology*, *50*, 147–161.

Rao, C., Kishan Prasad, H., Sajitha, K., Permi, H., & Shetty, J. (2016). Item analysis of multiple choice questions: assessing an assessment tool in medical students. *International Journal of Educational and Psychological Researches*, *2*(4), 201. https://doi.org/10.4103/2395-2296.189670

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests.* Chicago, US: The University of Chicago Press.

Regueras, L. M., Verdú, E., Munoz, M. F., Perez, M. A., de Castro, J. P., & Verdú, M. J. (2009). Effects of competitive e-learning tools on higher education students: a case study. *IEEE Transactions on Education*, *52*(2), 279–285. https://doi.org/10.1109/TE.2008.928198

Reise, S. P. (1990). A comparison of item- and person-fit methods of assessing model-data fit in IRT. *Applied Psychological Measurement*, *14*(2), 127–137. https://doi.org/10.1177/014662169001400202

Rizopoulos, D. (2006). ltm: an R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software*, *17*(5), 1–25. https://doi.org/10.18637/jss.v017.i05

Rizopoulos, D. (2017). Package “ ltm .” Retrieved from https://github.com/drizopoulos/ltm

Romero, G. M. O., Rojas, P. A. D., Domínguez, O. R. L., Pérez, S. M. P., & Sapsin, K. G. (2015). Difficulty and discrimination of the items of the exams of reasearch methodology and statistics. *Edumecentro*, *7*(2), 19–35.

Sinharay, S. (2003). *Bayesian item fit analysis for dichotomous Item response theory models*.

Thissen, D., & Wainer, H. (2001). *Test scoring*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Verhoeff, T. (1997). *The Role of competitions in education*. *Future world: Educating for the 21st century*. Retrieved from http://olympiads.win.tue.nl/ioi/ioi97/ffutwrld/competit.html

Zamri Khairani, A., & Shamsuddin, H. (2016). Assessing item difficulty and discrimination indices of teacher-developed multiple-choice tests. In S. Fun Tang & L. Logonnathan (Eds.), *Assessment for Learning Within and Beyond the Classroom* (pp. 417–426). Springer Science & Business Media. https://doi.org/10.1007/978-981-10-0908-2_6