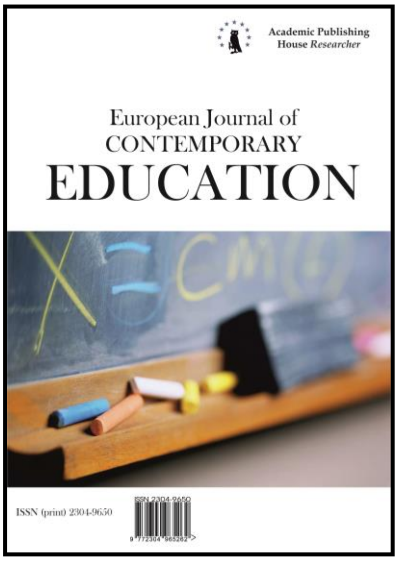
Published in the Slovak Republic
European Journal of Contemporary Education
E-ISSN 2305-6746
2018, 7(3): 448-457
DOI: 10.13187/ejced.2018.3.448
www.ejournal1.com



The Problems of Contemporary Education

Competitive Learning Using a Three-Parameter Logistic Model

Semei Coronado [a,*], Salvador Sandoval-Bravo [a], Pedro Luis Celso-Arellano [a], Ana Torres-Mata [a]

[a] Centro Universitario de Ciencias Económico Administrativas, Universidad de Guadalajara, Mexico

Abstract

The purpose of this paper is to analyze the test applied at the eighth Statistics II tournament to students from the University Center for Economic and Administrative Sciences of the University of Guadalajara, for the purpose of determining whether it promotes competitive learning among students. To achieve this, Item Response Theory (IRT) is used, specifically in the form of a three-parameter logistic model. The findings show that approximately 20 % of the participating students performed at a level ranging from outstanding to satisfactory, while the rest had a performance that fell between regular and poor. The findings also indicate that participating students were motivated by academic competition and the opportunity to improve their skills in the area of statistics. Moreover, we concluded that the tournament's assessment instruments need to be substantially improved in terms of design and the content of the items.

Keywords: competitive learning, Item response theory, Logic model.

1. Introduction

Meaningful learning techniques are intended to teach learners to solve problems or master certain topics and areas of knowledge (Hierro et al., 2014). Academic competitions, for their part, encourage better performance among students (Regueras et al., 2009). In this sense, Cantador & Conde (2010) and Lawrence (2004) conclude that skills and knowledge tournaments stimulate healthy and fair competition among students and generate higher implicit motivation in them.

Since the tournament constitutes an academic challenge, students require what is known as competitive learning (Johnson, Johnson, 2002; Kim, Sonnenwald, 2002; Owens, Straton, 1980),

* Corresponding author
E-mail addresses: semei.coronado@academicos.udg.mx (S. Coronado),
salvsanb@cucea.udg.mx (S. Sandoval-Bravo), pcelso@cucea.udg.mx (P. Luis Celso-Arellano),
anatorres@cucea.udg.mx (A. Torres-Mata)

which essentially consists of outperforming the other tournament competitors academically by obtaining a better result on the test. In this sense, competitions revolving around academic abilities and knowledge significantly improve participants' performance, through intellectual challenges and active experience, thus promoting confidence and motivation in the students who take part in the event, even those who are seen as weaker (Carpio Cañada et al., 2015; Fasli, Michalakopoulos, 2015; Lawrence, 2004; Verhoeff, 1997).

The Department of Quantitative Methods (DQM) of the University of Guadalajara (UdG) organizes an annual Statistics II Tournament (ST_II) for the purpose of promoting academic competition and, implicitly, competitive learning in the area of statistics among interested students enrolled at the University Center for Economic and Administrative Sciences (CUCEA). The event consists of two rounds, and the model examined in this paper refers to the test used in the first round.

Among the extensive literature that considers the application of tests of abilities and knowledge in university settings, the so-called Item Response Theory (IRT) can be found. This theory, among other things, serves to analyze curricular content, adequate item design, the recognition of teaching-learning problems and the identification of students with low or high academic ability (Awopeju, Afolabi, 2016; Balmori et al., 2011; DiBattista, Kurzawa, 2011; Ingale et al., 2017; Mitra et al., 2009; Rao et al., 2016; Romero et al., 2015).

In our opinion, there is no literature that analyzes multiple-choice tests used in academic competitions at the university level. For this reason the present study is relevant, as it examines this evaluation instrument using IRT theory to identify the type of students who participate in ST_II and determine whether the event fulfills the objectives proposed by the DQM, which include the promotion of competitive learning, the application of problem-solving skills and the general development of students' statistical abilities.

The paper is structured as follows: section 2 explains the methodology and data used; section 3 shows the results; finally, section 4 presents conclusions.

2. Data and methodology

Data

The Statistics II course is taught at CUCEA in eleven of the thirteen undergraduate programs offered at that particular university campus. The DQM has organized ST_II every year since 2009, with the intention of enhancing the statistical abilities of the university's graduates by having them solve practical problems that require them to apply what they have learned during the course, while promoting academic competition among the students (DMC, 2017).

Statistics courses are taught by a group of specialized professors that make up the Academy of Statistics, which belongs to the DQM. A committee of professors from this Academy is in charge of organizing the event, both the logistics and the academic aspects, including content selection and the design and development of test questions. The 2017 edition of ST_II consisted of two rounds; the first round was open to interested students who were taking the Statistics II course. The test applied in this first round consisted of 20 multiple-choice questions. The test is not included in this paper for reasons of confidentiality, as requested by the DQM.

The test corresponding to the first round was taken by 99 students of an eligible population of 1,666, which represents 5.94 % of the total. The finalists of this first round were the 20 students obtaining the highest score on the multiple-choice test. These students were chosen for the second round, and the students with the highest scores in this second round were awarded prizes such as scholarships, university books, graphing calculators and financial calculators sponsored by different organizations.

The test examined in this paper was the one applied in the first round of the ST_II; considering the size of the population, the sample size is quite acceptable. As mentioned above, the test consists of 20 multiple-choice questions, each with five options: one right answer and four distractors. The contents included sampling theory, parameter estimation, and hypothesis testing for large and small samples. For more information about this test, go to http://metodos.cucea.udg.mx/estadistica.php.

Multiple-choice tests, like all evaluation instruments, have advantages and disadvantages. The advantages include the evaluation of critical comprehension and knowledge and the memorization of simple concepts; they can assess whether algorithms and procedures are performed accurately in solving problems or in ordinary calculations; they reduce the probability of guessing right (probability declines as the number of distractors rises). When it comes to disadvantages, syntax errors are the most frequent, as they generate confusion and misinterpretation both in the formulation of questions and in the answers and distractors (Best, Kahn, 2006; DiBattista, Kurzawa, 2011; Miller et al., 2009; Zamri Khairani, Shamsuddin, 2016).

Methodology

In educational research, one of the fundamental purposes is to quantify variables that will estimate the performance achieved in the teaching-learning process. This variable is known as a latent or treatment variable, and it quantifies a non-observable underlying characteristic. In our case, it will measure the ability level of ST_II contest participants in the Statistics 2 course. To analyze this variable, IRT (Baker, Kim, 2017) is used to analyze the magnitude and characteristic values of this latent variable.

Different papers mention the advantages of using IRT. These advantages include the following: greater emphasis can be placed on the distinctive characteristics of the questions rather than on the essential properties of the tests; tests can be modeled in a non-linear way, with one, two or even three parameters, in order to establish with greater certainty which model lends itself best to the distribution of the available data; the scale of the latent variable (ability level) is in the $(-\infty,\infty)$ interval, although it can easily be transformed into another scale; it has the quality of being invariant; and finally, the values of the parameters calculated for the questions and the participating students are independent with respect to the sample used (Aiken, 1979, 2003; Finch, French, 2015; Furr, Bacharach, 2013; Hambleton et al., 1991; Hambleton, Jones, 1993; Muñiz, 2010; Zamri Khairani, Shamsuddin, 2016).

Taking this into consideration, we used IRT to study the test corresponding to the first round of ST_II, specifically with the Rasch logistic model (RM), developed by Rasch in 1980. An important quality of IRT models is that they show the relationship between the variable of interest (in our case the latent variable that measures the students' ability to do statistics) and the probability of answering a certain item right, which can be represented with RM. The RM models operate under three basic assumptions: 1) the function that relates the latent variable and the probability of answering the question right is monotonous and increasing, 2) there is only one latent variable, and this one feature is measured by the entirety of the questions on the test, and 3) there is no correlation between the results of the questions, i.e., the latent variable is locally controlled and independent of each question(Finch, French, 2015).

RM models require the binary codification of each answer to a question (obviously, 1 for right and zero for wrong). Our research looks at three models: 1) the one-parameter logistic model (1PLM), 2) the two-parameter logistic model (2PLM), and 3) the three-parameter logistic model (3PLM). All three models, 1PLM, 2PLM, and 3PLM, can be defined as a three-parameter logistic model (3PLM) (Ark et al., 2016).

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

where $P\left(x_{ij}=1\middle|\theta_j,a_i,b_i,c_i\right)$ is defined as the probability of student j answering right (1), as opposed to the alternative of answering question i wrong (0); a_i represents the slope, given the curvature of the model used; b_i indicates the difficulty of the question; c_i represents the possibility of guessing the right answer to question i; and finally θ_j indicates the ability shown by student j.

3. Results

The ST_II test consisted of twenty multiple-choice questions, with 5 possible alternatives per item; one of them was the right option and the remaining four were distractors. Figure 1 below shows the percentages of right answers compared to wrong answers given by the students that participated in the first round.

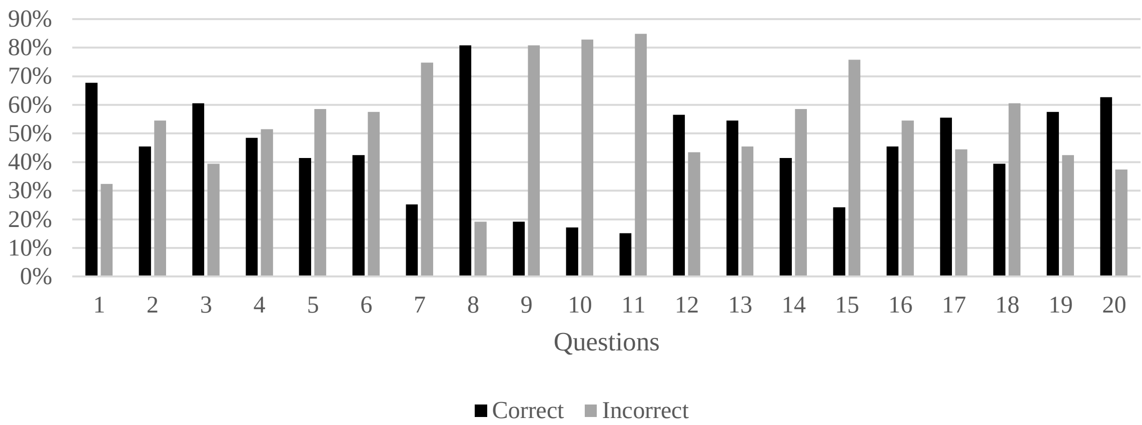


Fig. 1. Percentage of right and wrong answers per test item
Note: prepared by the author based on the results of R

Figure 1 shows a decline in the percentage of right answers from item 1 to 11; after the halfway point, only 4 questions had a higher percentage of wrong answers than of right answers. In summary, 45 % of the items were answered right and 55 % wrong.

Table 1 shows the descriptive statistics results of the applied test. It shows positive asymmetry, i.e., the number of participants with poor scores exceeds the number of participants with good scores.

Table 1. Descriptive statistics

Mean	9.03
Median	8
Mode	6
Standard deviation	3.86
Sample variance	14.89
Kurtosis	-0.73
Asymmetry	0.48
Minimum	3
Maximum	19

Note: prepared by the author based on the results of R

The participants' scores ranged from 3 to 19 right answers, i.e., their grades ranged between 15 and 95 on a scale of 0 to 100. The 20 participants with the highest scores were chosen for the second round, and their scores ranged between 13 and 19 right answers (grades between 65 and 95). The students selected for the second round made up 20.20 % of the total, while 79.80 % were disqualified. Figure 2 shows the percentages of grades obtained by the students who passed to the second round.

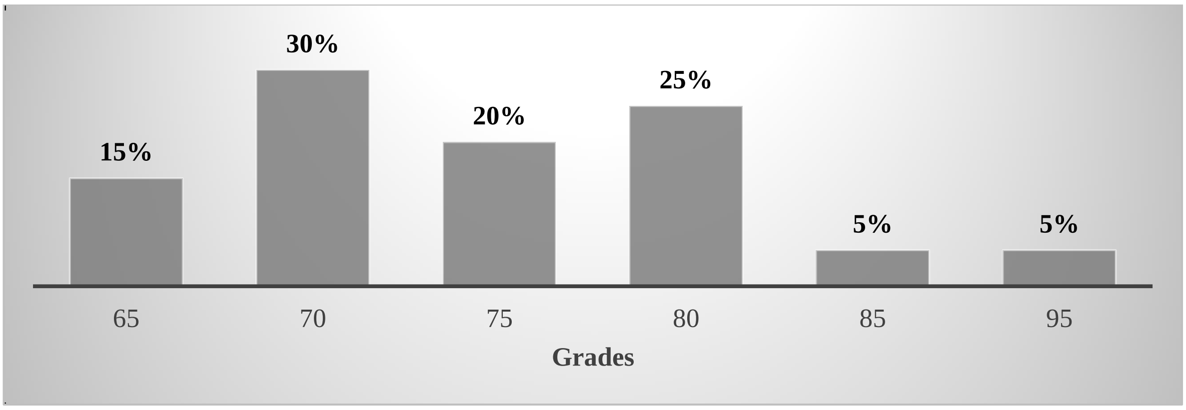


Fig. 2. Percentages of grades obtained by students who passed to the second round
Note: prepared by the author based on the results of R

There are several criteria to select the model (1PLM, 2PLM, 3PLM) with the best fit for the data; for example, the Aikaike (AIC) criterion, the Bayesian criterion (BIC), the likelihood ratio test (LRT), Relative Efficiency (RE), a latent variable simulation through the estimation of the Kernel density function, the information function test, and the goodness-of-fit test. Each of these criteria was applied to determine the model that best fit the data; the results are available from the authors upon request. The criterion that was selected for the model was the BIC, since it suits our purposes better than the Aikaike criterion (AIC). However, there are other criteria that measure the goodness-of-fit of the different models, and a definitive conclusion cannot be reached as to which of them is the best (Finch, French, 2015).

Table 2 shows the statistical information that allowed us to choose the model that best fit the set of available data. It is worth mentioning that 3PLM has a higher BIC statistic than 1PLM, and an AIC below that of 1PLM. It also had better behavior in the results of the aforementioned decision criteria, and on this basis it was decided to apply 3PLM. The information was processed using Latent Trait Models under IRT software (Rizopoulos, 2017), in addition to the free-use statistical package R.

Table 2. AIC and BIC statistical values for the ST_II test

Model	AIC	BIC
1PLM	2382.06	2436.56
2PLM	2360.90	2464.70
3PLM	2348.87	2455.27

Note: developed by the author based on the R results

The results of the 3PLM coefficients are shown in Table 3, ordered from the lowest to the highest level of difficulty. The results for the test's b_i coefficients are ordered from the easiest question (question 3) to the most difficult (question 15). This information coincides with the percentages of right and wrong answers shown in Figure 1. The b_i coefficients can have both positive and negative values; thus, the values close to zero represent questions of moderate difficulty; negative values indicate relatively easy items (below average), while positive values indicate relatively difficult items (above average). However, it can be concluded that the test has a high to average level of difficulty.

The fourth column of Table 3 represents the probability of an average student answering question *i* right. We can see that this value increases as the question's level of difficulty decreases (Rizopoulos, 2006).

Table 3. 3PLM coefficients and Answer Probability

Items	c_i	b_i	$P(x=1\|z=0)$
3	0.02	-0.33	0.67
12	0.00	-0.21	0.61
4	0.00	0.05	0.48
20	0.30	0.12	0.61
17	0.23	0.26	0.51
19	0.27	0.26	0.54
13	0.26	0.39	0.49
8	0.71	0.55	0.78
18	0.12	0.65	0.30
2	0.22	0.68	0.37
14	0.24	0.95	0.33
16	0.32	1.09	0.38
1	0.60	1.16	0.63
6	0.30	1.22	0.35
7	0.12	1.36	0.17
5	0.35	1.68	0.37
10	0.12	1.99	0.13
11	0.12	2.27	0.12
9	0.17	2.62	0.18
15	0.23	3.04	0.24

Note: prepared by the author based on the results of R

The discrimination coefficient (parameter a) that was obtained from the model was 2.10 for each of them, which indicates that the characteristic curve for each of the items has a steep slope, as detailed below.

Table 3 shows that the item with the highest c_i coefficient is item 8, with a b_i value of 0.55 and a P value of 0.78. On the other hand, there are two items with a c_i coefficient equal to zero, items 12 and 4; another item near zero is item 3, with a c_i value of 0.02. According to Figure 1, these items were answered wrong by slightly more than 40 % of the students, and item 3 by slightly less than 40 % of the students; however, these are easy items according to the b_i coefficient. Items 7, 9, 10, 11 and 15 had the highest percentages of wrong answers; however, they are relatively difficult items according to the b_i coefficient of 3PLM. A review of the individual scores shows that they were answered right by all the students who advanced to the second round of the tournament.

Figure 3 shows the characteristic curves for each ST_II test question, ordered from lowest to highest level of difficulty. We can observe that the ICC of item 8 shows that students with average ability have a nearly 74 % chance of answering it right. The same goes for item 1, where students with below-average ability have a 60 % chance of answering it right. These graphs show that 90 % of the items on the ST_II test are relatively easy.

The application of the entire information test in the (-10, 10) interval (Baker, 2001; Rizopoulos, 2017) yielded an information total of 25.04; the same test in a (0,10) interval yielded an information total of 20.17, which equals 80.55 %, implying that 19.45 % of the students have an ability level below zero. This behavioral pattern made evident by the test can be seen in Figure 4, in which the total information curve has an approximately symmetrical pattern, skewed slightly toward the left. In this sense, we can conclude that the ST_II test is actually aimed at the highest-performing students of CUCEA in the area of statistics, i.e., those attracted to a challenging and motivating academic competition; this leads to the conclusions that the ST_II adequately meets the objectives proposed by the DQM in terms of actively promoting meaningful and competitive learning among students.

We conducted a Kernel Density Function test to calculate variable θ (ability) for our data set, which resulted in an estimated value tending toward zero, with an asymmetrical behavior skewing positive, which closely resembles the behavior observed on the total information curve in Figure 4.

The calculated ability level of the 20 students selected for the second round ranged between 0.72 and 3.33, and the rest of the participants scored between -2.00 and 0.47. Of the latter group, 65.67 % have below-average ability, which suggests that the ST_II test does select the most capable students for the second round of the contest.

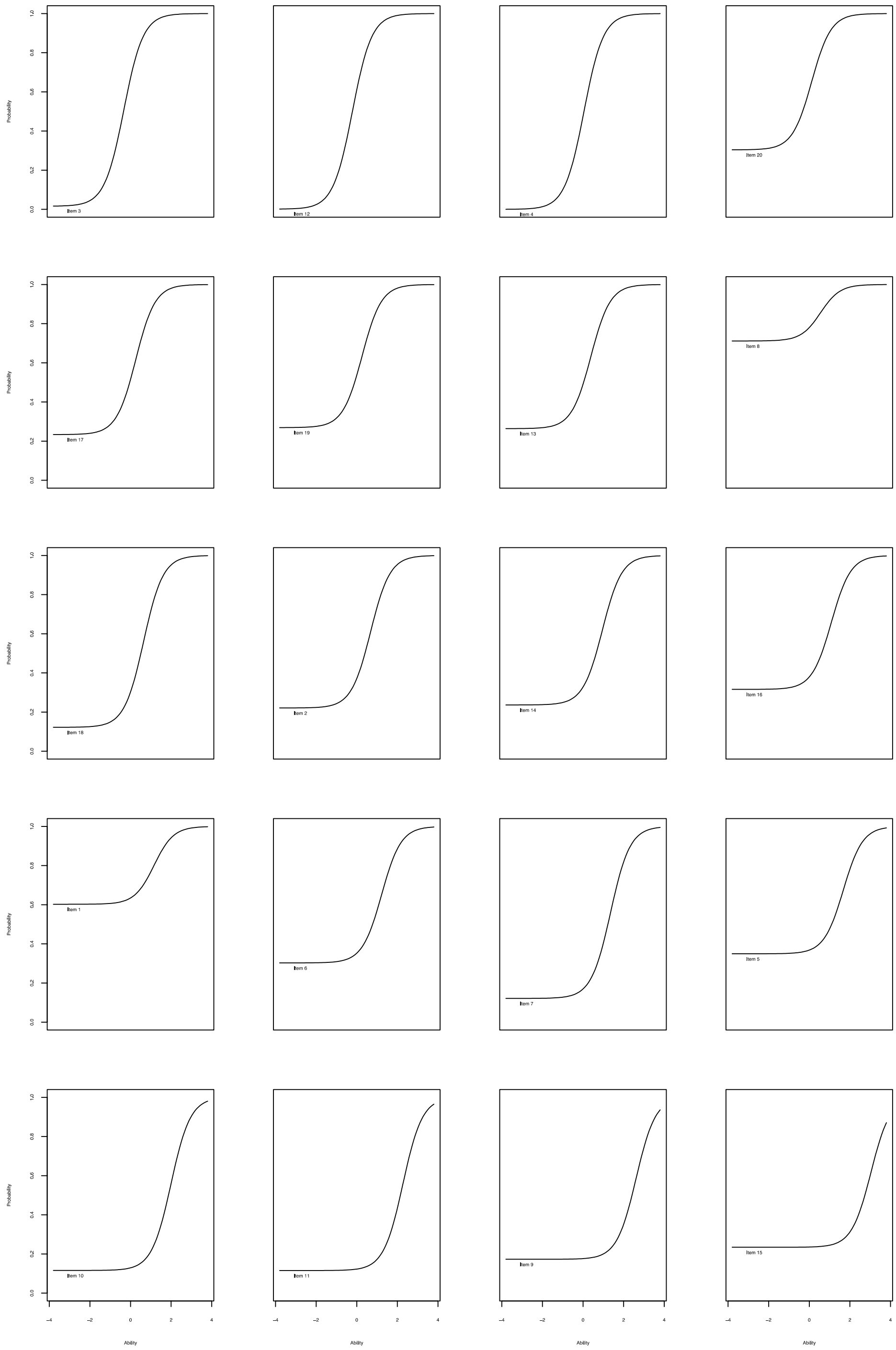


Fig. 3. Characteristic curves by question
Note: prepared by the author based on the results of R

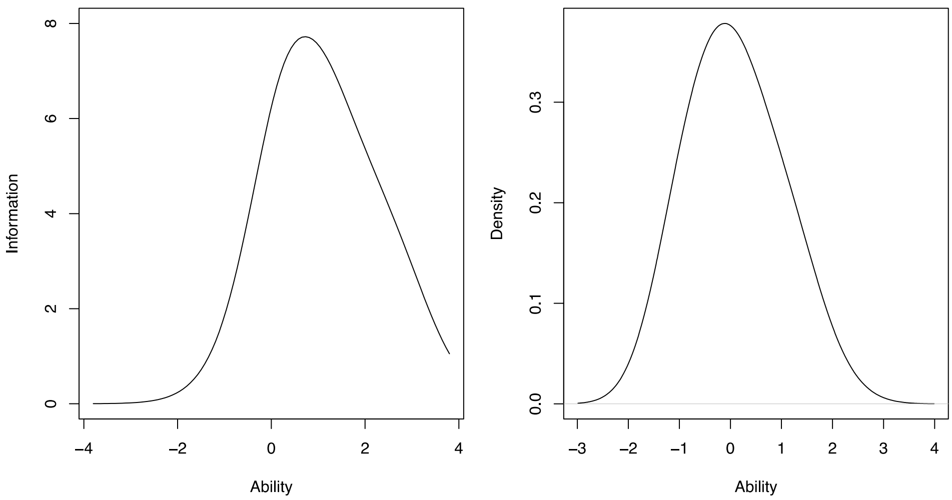


Fig. 4. Test Information Function and Kernel Density (left and right)
Note: prepared by the author based on the results of R

4. Conclusion

This research is the first study to use IRT to look at the application of a multiple-choice test in a university-level statistics competition that promotes active and competitive learning among students. However, a literature review finds similar studies using IRT for partial and departmental exams from a wide range of undergraduate courses (Awopeju, Afolabi, 2016; Balmori et al., 2011; DiBattista, Kurzawa, 2011; Escudero et al., 2000; Gajjar et al., 2014; Ingale et al., 2017; Marie, Edannur, 2015; Mitra et al., 2009; Rao et al., 2016; Romero et al., 2015).

Given the results of our research, we can conclude that the exam taken by students in the ST_II competition is designed for well-prepared students with above-average abilities for statistics. The students chosen for the final round achieved grades between 65 and 95. IRT theory distinguishes between the different types of ability that students show on a particular test, while also considering whether the question options (distractors and right answers) were well designed (McDonald, 2017) since they can significantly influence the results and conclusions derived from the data (DiBattista, Kurzawa, 2011).

The ST_II competition is an attempt by the DQM to promote active and competitive learning, as well as to provide high-level academic challenges for CUCEA students. Student participation fell below expectations, since only a small portion of the eligible students took part. However, we were able to show that those who did participate constitute a small core of students who are genuinely interested in competing intellectually in statistics, and are implicitly motivated by the course they are taking.

Furthermore, we suggest that the professors who designed the ST_II test undertake ongoing training in the design of questions that assess academic performance, since our research showed that most of the items are similar to the problems and exercises found in statistics textbooks, which are not necessarily ideal for this type of event. We also suggest a general evaluation of the teaching activities to ensure that they are aligned with the objectives and contents of the ST_II: this will help to devise a robust and well-designed set of questions for future events (Zamri Khairani, Shamsuddin, 2016).

5. Acknowledgments

We would like to thank the DQM and the Academy of Statistics at CUCEA for providing the information needed to conduct this research.

References

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

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

Ark et al., 2016 – *Ark, L. A., Bolt, D. M., Wang, W., Douglas, J. A., & Wiberg, M.* (2016). *Quantitative Psychology* (Vol. 167). Retrieved from: https://doi.org/10.1007/978-3-319-56294-0

Awopeju, 2016 – *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. Retrieved from: https://doi.org/10.19044/esj.2016.v12n28p263

Baker, 2001 – *Baker, F. B.* (2001). The basics of item response theory. Evaluation. ERIC. Retrieved from: https://doi.org/10.1111/j.1365-2702.2011.03893.x

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

Balmori et al., 2011 – *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, Kahn, 2006 – *Best, J.W., Kahn, J.V.* (2006). Research in education. Pearson (10th ed.). Boston, MA: Pearson Education.

Cantador, Conde, 2010 – *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 et al., 2015 – *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. Retrieved from: https://doi.org/10.1111/jcal.12075

DiBattista, Kurzawa, 2011 - *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. Retrieved from: https://doi.org/10.5206/cjsotl-rcacea.2011.2.4

Escudero et al., 2000 – *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, Michalakopoulos, 2015 – *Fasli, M., Michalakopoulos, M.* (2015). Learning through game-like simulations. *Innovation in Teaching and Learning in Information and Computer Sciences*, 5(2), 1–11. Retrieved from: https://doi.org/10.11120/ital.2006.05020005

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

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

Gajjar et., al 2014 – *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. Retrieved from: https://doi.org/10.4103/0970-0218.126347

Hambleton et al., 1991 – *Hambleton, R.K., Swaminathan, H., Rogers, H.J.* (1991). Fundamentals of item response theory. Retrieved from: https://doi.org/10.2307/2075521

Hierro et al., 2014 – *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. Retrieved from: https://doi.org/10.4995/redu.2014.5634

Ingale et al., 2017 – *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. Retrieved from: https://doi.org/10.18203/2394-6040.ijcmph20171764

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

Hambleton, Jones, 1993 – *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. Retrieved from: https://doi.org/10.1097/01.mlr.0000245426.10853.30

Kim, Sonnenwald, 2002 – *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. Retrieved from: https://doi.org/10.1002/meet.1450390107

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

Marie, Edannur, 2015 – *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, 2017 – *McDonald, M.* (2017). The nurse educators guide to assessing learning outcomes. Burlington,MA: Jones & Bartlett Learning.

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

Mitra et al., 2009 – *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, 2010 – *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, Straton, 1980 – *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 et al., 2016 – *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. Retrieved from: https://doi.org/10.4103/2395-2296.189670

Regueras et al., 2009 – *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. Retrieved from: https://doi.org/10.1109/TE.2008.928198

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

Rizopoulos, 2017 – *Rizopoulos, D.* (2017). Package 'ltm'. Retrieved from: https://github.com/drizopoulos/ltm

Romero et al., 2015 – *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.

Verhoeff, 1997 – *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, Shamsuddin, 2016 – *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. Retrieved from: https://doi.org/10.1007/978-981-10-0908-2_6