Promoting Equality in Higher Education: Development and Internal Validity of a Selection Test for Science University Degrees in Ecuador

Iván Sandoval[a], Raquel Gilar-Corbí[b], Alejandro Veas[b] & Juan-Luis Castejón[b]

[a] National Polytechnic School of Ecuador. Department of Basic Training.

[b] University of Alicante. Department of Developmental Psychology and Didactics.

Abstract:
The use of tests or examinations for student admission is an extended strategy in most higher education institutions. However, tests in Latin-American countries have not received attention regarding the validity and interpretation of these measures. As such, the aim of this study is to analyse the psychometric properties of a new university entrance examination test in Ecuador in a community sample of 1238 university students (28.10 % female). A two-parameter multidimensional item response theory (MIRT) model was used to calibrate item difficulty and discrimination parameters as well as differential item functioning (DIF) by gender. The final instrument was composed by 71 items, which was considered appropriate to ensure the measurement precision of all levels of students' achievements.

Keywords:
University selection test, polytechnic school, Test Validation, multidimensional item response theory

1 Introduction

The evaluation process is a fundamental tool for training and measuring the impact of educational systems. During the past decades, different studies have attempted to address the quality of measuring external performance assessment tests in international contexts, such as the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science (TIMSS), among others (Carnoy et al., 2015; Haladyna & Downing, 2004; Kane, 2013). Moreover, within the fields of science, technology, engineering, and mathematics (STEM), these tests have been considered as major indicators for ranking countries in international assessments (Liou & Bulut, 2020), including education policy initiatives from different countries (Ho, 2016; Lietz & Tobin, 2016; Schmidt & Burroughs, 2016).

The use and application of achievement measures are also crucial in higher education, especially in the tests or examinations designed for student admission and in ensuring the comparability of results (Coe, 2010). In this context, a number of European studies have considered the theoretical and validity measurements of certificate examinations (Baird et al., 2000; Coe 2007; He et al., 2018) or university entrance examinations (Veas et al., 2020a 2020b). However, tests in Latin-American countries have received no attention regarding the validity and interpretation of these types of measures, considering that more social disadvantages exist among students.

This study aims to fill this gap by considering the social and academic selection context of a large polytechnic school in Ecuador to develop and validate a new measurement instrument that assesses the science and language level of students who want to enrol into a STEM degree. This study investigates the test functioning of a new measure under a multidimensional item response theory (MIRT) model, which includes a range of item difficulty and discrimination parameters along a latent construct. Moreover, differential item functioning (DIF) is also explored to determine item gender bias, which may affect the measurement precision of the instrument.

1.1 Social diversity in the higher education selection process

In the Latin-American and Caribbean regions, student retention and dropout are distinct and negative realities in all levels of education. Traditionally, possible reasons focused on the students' characteristics, although these reasons finally moved on from the relation between the students and the institutions to the institutions' responsibility to address a massive and heterogeneous group of students (Braxton et al., 1997; Himmel, 2002).

The negative impact of inequality on student achievement across cultures has proved to be consistent in multiple studies (Alexander et al., 2001; Georges & Pallas, 2010; Ma et al., 2018). The reduction of inequalities in the access to and completion of higher education has been a strategic target in different countries (Eurydice, 2012). This priority is given by the fact that the probability of university enrolment and retention differ substantially across social backgrounds. Important theoretical models have been proposed to provide a consistent explanation of this phenomenon. For instance, Bourdieu and Passeron (1990) referred to the theory of cultural reproduction, where children in the highest classes have advantages in gaining educational credentials due to their possession of cultural capital. Breen and Golthorpe (1997) conceived of the educational success of students in terms of evaluating costs and benefits and the perceived probability of success outcomes. Another recent view is the consideration of learning capitals from the actiotope model of giftedness (Ziegler & Baker, 2013; Ziegler, Chandler, et al., 2017; Ziegler et al., 2019). This model can be extended to all students' levels of achievement, as it endorses different factors in a dynamic complex system (Ziegler, Balestrini, et al., 2017; Ziegler et al., 2013; Ziegler & Stoeger, 2017). Considering educational capital as external influence of self, inequality may affect to both cultural and social capitals. In this context, it is clear that performance differences in the transition to higher education are also produced by social selection during the earlier stages of schooling.

1.1.1 Selection process in Ecuadorian higher education institutions

Special efforts have been taken in Latin-American higher education institutions for them to be considered as relevant actors of social development (Arocena & Sutz, 2005). During the twentieth century, important student movements triggered the so-called University Reform Movement ([URM]: Ribeiro 1971; Tünnermann, 2000), allowing the inclusion of social policies based on the increasing enrolment in higher education despite political or military controversies over the past decades. This spread of democracy in the higher education system has met the goals of stronger teaching and research standards (de Moura Castro & Levy, 2001).

According to social demands, polytechnic schools began to provide qualified professional techniques in fundamental areas of progress in the country. Starting from religious institutions, expert groups determined objective criteria for the development of undergraduate degrees and assessment standards (Contreras & Maluk 2017). For example, polytechnic schools, in comparison to universities, should provide at least 70 % of the professional titles in the basic and applied sciences with the guarantee of excellence and academic rigor among the scientific field. As a more formal example, the National Polytechnic School of Ecuador (NPSE), created on the 30th of August, 1986, is a public higher education institution that is in line with a mass access model policy. As such, and in line with the National Secretariat of Higher Education, Science, Technology, and Innovation (SENESCYT), the student admission process considers vulnerable groups of students from the perspective of social inequality. During this process, an applicant of an Ecuadorian higher education institution must meet certain requirements, such as taking the National Exam of Educational Evaluation (*Ser Bachiller*) and completing the Associated Factors Survey. Although there is no minimum score needed to apply for a degree (it varies depending on the institution), the allocation of places is automatically made according to the application score, the availability of places in each institution, and the demand that exists for a degree in a given period. In short, applicants with the highest scores in the exam are more likely to get a place (SENESCYT 2018). However, in 2014, SENESCYT implemented a positive action policy that would expand access to higher education for socially and economically vulnerable applicants through the Affirmative Action Programme. An applicant who has been included in this programme can preferentially apply to 15 % of the academic places offered by public higher education institutions, even if their application score has not met the minimum required for a specific institution during the admission process (Di Caudo, 2015).

The determination of the beneficiaries of the Affirmative Action Programme is achieved through the analysis of each applicant's self-declared information in the Associated Factors Survey. Then, a vulnerability index is calculated, and the lowest values correspond to applicants from traditionally-excluded groups, those who have a disability, or those who are placed in the lowest decile according to their socio-economic status. In this sense, applicants in situations of greater vulnerability are usually assigned to the Affirmative Action population segment.

Since 2017, SENESCYT has included, among the new students entering the levelling course of the NPSE, those from the Affirmative Action population segment. These students' average application scores were observed to be lower than those obtained by other population segments, denoting poor previous academic preparation. During the first year of study, students from vulnerable groups generally show lower academic performance when compared to that of their peers from other population segments. Additionally, during high school, mathematics typically has the lowest indicator as it is perceived as being more complicated.

1.2 Multidimensional item response modelling in test development

Research has begun to apply item response theory models (IRT) to validate measures in different fields (Christensen et al., 2019) with a general use in educational assessment (Embretson, 1984; Hartig & Höhler, 2009), and more specifically, in science education (Kaspersen & Ytterhaug, 2020). Van der Linden (2017) claimed that IRT analysis, which focuses on the quality of items when measuring underlying constructs, perfectly complements classical test theory approaches. In this sense, IRT models has become popular in test construction, including large-scale educational assessment, to optimize item selection and scale validation across diverse populations (Khorramdel & von Davier, 2016). Differing from classical test theory, which considers that an observed test score is composed by a true score and a random component, IRT considers that the probability of a person's expected response to an item is a mathematical function of that person's ability and one or more parameters that characterize the item (Reckase, 2009).

MIRT models have appeared in the research literature since the 1980s (e.g. Bock & Aitken 1981). The purpose of MIRT is to provide a model with an appropriate representation of data (given an incidental vector, θ, which describe the locations of individuals), and structural parameters are used to describe the functioning of the test items in a m-dimensional space, where m is the number of dimensions used to model the data (Reckase, 2009).

MIRT models for dichotomous items (those with two categories) are one of the most important in achievement tests. A multidimensional extension of the two-parameter logistic model is given by the following equation (Reckase, 2009):

$$P\left(U_{ij} = 1 \middle| \theta_j, a_i, d_i\right) = \frac{e^{a_{i\theta'}{}_{j+d_i}}}{1 + e^{a_{i\theta'}{}_{j+d^i}}}$$

where U_{ij} is the score for person *j* in item *I*; θ_j is the parameter that describes the ability of the *j*th person in item *I*; and a_i is a parameter related to the maximum slope of the item characteristic curve along the latent construct which measures the item's discriminating power. Given the multidimensional nature of the model, a slope/intercept form, $a\theta + d$, is introduced in the equation, where *d* is the result of the *ab* item interaction (*b* = item difficulty). Therefore, *a* is a 1 × *m* vector of the item's discrimination parameters and θ is a 1 × *m* vector of the person's coordinates, with *m* indicating the number of dimensions in the coordinate's space.

The present study

The NPSE has two levelling courses: one for the engineering, sciences, and administrative sciences and one for the superior technological level. The aim of the levelling course, beyond its academic purposes, is to enrol new students to the university context until the study programme's completion. However, according to the information provided by the management and information processes department, approximately 40 % of students who receive a score that is less than or equal to six points in all the subjects of the levelling course (mathematics fundamentals, geometry and trigonometry, physics, chemistry, and language and communication) during the first bimester abandon this course.

The antecedents described are of relevance to the Latin-American science university system, as it is crucial to determine the correct application of selection tests to students who may enrol in the levelling course. Otherwise, possible defects are associated with potential higher education dropout rates. For these reasons, the present study aimed to analyse the psychometric properties of a new selection test under the two-parameter MIRT model. The specific objectives were as follows: (1) to analyse model-data fit of items of the test, (2) measure the precision of items according to individual levels of ability, and (3) invariance properties according to gender.

2 Method

2.1 Participants

The sample comprised of 1238 newly enrolled students (*mean age* = 18. 85, *SD*= 1.84) from the NPSE for the first semester of 2019 (890 males, 348 females). Of these, 955 students (280 males, 675 females) were enrolled in the engineering levelling course, and 288 students (215 males, 68 females) were enrolled in the technology levelling course.

2.2 Measures

The design process of the test began with an initial survey administered to the professors assigned to the levelling courses. These professors identified the elemental topics that students usually present academic difficulties in. These topics were compared to those studied at the higher level of basic general education and the Baccalaureate, and a list of curricular content was elaborated upon to determine students' previous knowledge in order to receive an appropriate score in the levelling courses. Therefore, a diagnosis test was designed following this content criteria.

The first pilot test composed of 65 multiple choice items that evaluated the following topics: real numbers operations, polynomials operations, factoring system, equations and inequations, and functions in real geometry and trigonometry fields. After its first application in October 2018, the evaluation committee decided that the test should include items on language and communication, with a maximum length of 80 items. Therefore, the instrument was composed of a mathematics section (55 items) and a language and communication section (25 items). All of the items presented four alternative options, of which only one was correct.

2.3 Procedure

The data were collected on the 28th of March, 2019 in paper-pencil form from different classes in the NPSE. Before the application of the test, the university made an official announcement to students via email and provided an instructive link in the university webpage. Furthermore, professors in the department of basic training were recruited for responsibility over controlling the students during the test application as well as distributing and collecting the material. The duration of the test was two hours in accordance with the evaluation criteria.

Students' responses were sent to the admission and registration unit for computer correction. Every correct response was scored as 1 point, whereas incorrect responses were scored as 0 points. Incorrect responses were not penalised. Punctuation in each section was computed as 50 % of the total score.

2.4 Data analysis

Data analysis comprised of two phases of validation. First, the fitting quality of each item considering the multidimensional structure was analysed through the expected values of infit (weighted) and outfit (unweighted) mean square error, and statistics were determined to be between -2 and +2 according to standard criteria (Fan 1998). Second, regarding the content, this study employed the fitting quality of each item considering the multidimensional structure of the instrument. Next, with respect to the generalisability aspect of validity, this study conducted differential item functioning (DIF: Holland & Wainer, 1993) analysis among gender. A difference of 0.5 logits in the overall item difficulty across groups was considered as a substantial DIF. The mean item parameters were set to be equal over groups so that the differences in the parameter estimates could be directly compared. Moreover, the item discrimination index was analysed and a good index criterion was considered to be above 0 and below 2 (De Ayala, 2009). The parameters were estimated using the computer programme Conquest Version 2 (Wu et al., 2007) via the maximum likelihood method.

3 Results

The two dimensions of the entrance examination test were calibrated simultaneously. Table 1 shows the difficulty estimates, fit statistics, discrimination estimates, and DIF magnitudes for each item. All of the items showed excellent infit and outfit values, and most of them were close to 1.00. The discrimination parameters showed acceptable values in all of the items that belonged to the mathematics subscale. With respect to the language and communication subscale, items 56 (-0.33), 57 (-0.20), 64 (-0.07), 74 (-0.07), 75 (-0.04), 76 (-0.16), and 77 (-0.15) showed values below 0. Therefore, these items did not have enough power to discriminate between the more and less able students on the language and communication subscale.

DIF analyses were conducted to assess the model-data fit across gender. As indicated in Table 1, items 31, 56, and 61 showed significant DIF, implying that these three items were more difficult for females. The largest values were for items 56 and 61, which were included in the language and communication subscale.

When applying the purification procedure (Lord, 1980), items with non-adequate discrimination parameters or no DIF values were removed, and a new item parameter was implemented which considered 71 items: 54 belonged to the mathematics subscale and 17 belonged to the language and communication subscale.

An item-person map is provided in Figure 1, as it is possible to calibrate a person's measurement from low to high and item difficulty from easy to hard along the same latent trait scale. The two continuums on the left side of the figure indicate students' measures in the two dimensions of the test. Individuals who had high scores are placed at the top of the continuum and those who had lower scores are placed at the bottom. Moreover, the items that fall into each of the two dimensions are clustered on the right side. All of the items are distributed reasonably well along the latent construct. The students located at the medium side of the scale were targeted by the majority of items. The most difficult items were 59, 60, and 64, which belonged to the language and communication subscale; the easiest items were 7, 42, and 36, which belonged to the mathematics subscale. These items targeted an important proportion of low-ability students.

4 Discussion

This study aimed to analyse the psychometric properties of a newly developed version of a university entrance examination test in a sample of university students enrolled in the ESPN, one of the largest public institutions in Ecuador. This instrument was intended to ensure that students of a minimum curriculum level had access to the levelling courses under a global access policy that focuses on the population's social diversity and vulnerability.

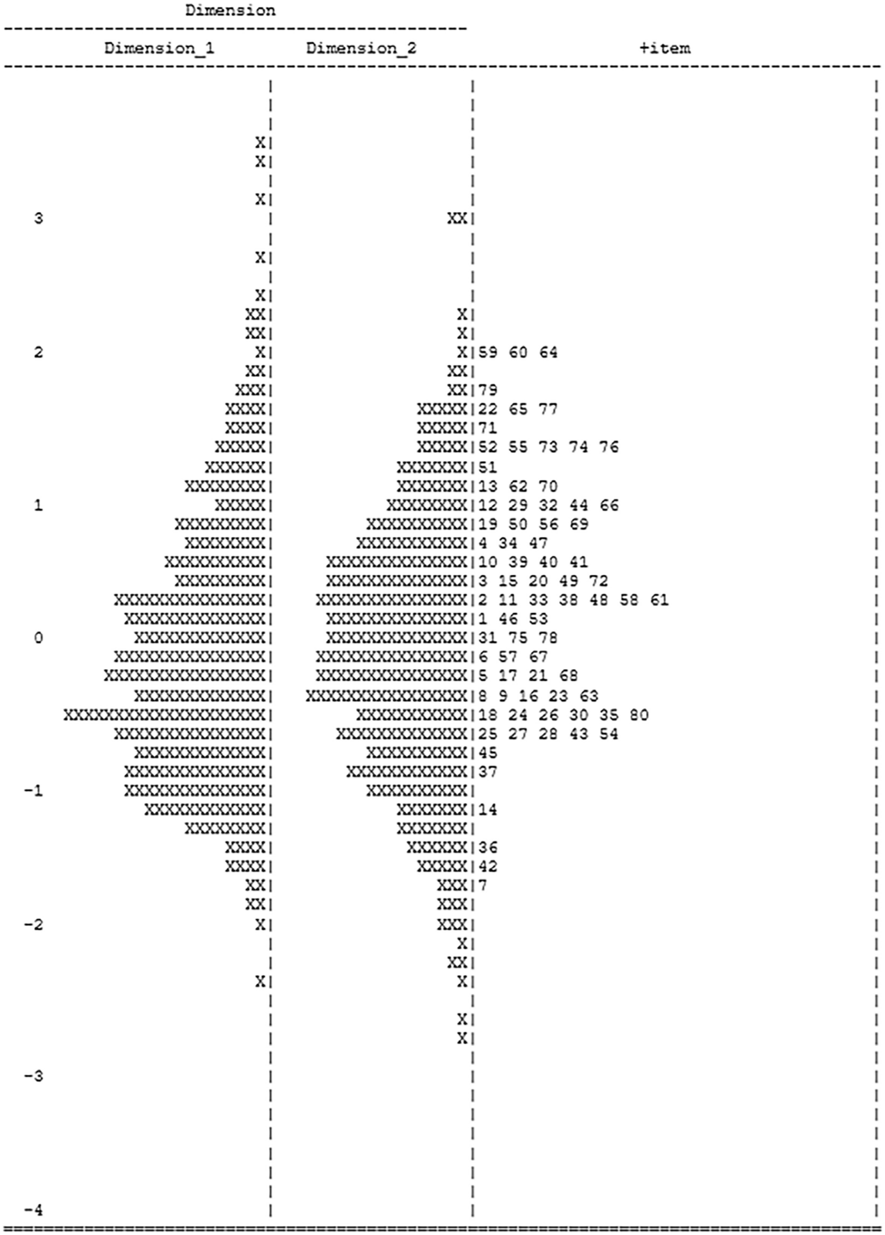
To gain a deeper understanding of the measurement precision, a two-parameter MIRT was implemented. Considering the initial 80 items distributed in two subscales (mathematics and language and communication), the results showed excellent item fit values. Nine items showed poor discrimination parameters or DIF; without these items, the new estimation provided acceptable values for all parameters. In general terms, the mathematic subscale showed better parameter values than the language and communication subscale. The item-person map showed that item difficulty was reasonably spread at the top of the map, as the main objective of the scale was to detect all students' achievement levels using adequate measurement precision.

By using construct validation measures, it is possible to extend appropriate measurement practices in Latin-American universities and ensure equality processes through innovative network policies among institutions (Arocena & Sutz, 2001). Because of the widespread concern over the social needs of

Items parameters

Items	Item difficulty (SE)	Infit	Outfit	Item discrimination	Gender DIF
1	0.18(0.07)	1.00	1.00	0.89	0.21
2	0.29(0.08)	0.99	1.04	1.29	0.15
3	0.45(0.07)	0.99	0.99	1.17	0.09
4	0.76(0.07)	0.99	1.01	0.75	0.04
5	-0.24(0.07)	1.00	1.01	0.95	0.18
6	-0.04(0.08)	0.99	1.02	1.47	0.00
7	-1.61(0.12)	1.00	1.01	1.85	0.03
8	-0.36(0.07)	1.00	0.98	1.02	0.29
9	-0.28(0.07)	0.99	0.97	1.03	0.06
10	0.63(0.09)	0.99	1.07	1.88	0.13
11	0.36(0.06)	1.00	1.00	0.48	0.05
12	1.03(0.07)	0.99	1.02	0.82	0.06
13	1.29(0.08)	0.99	1.03	0.98	0.10
14	-1.18(0.08)	1.01	0.98	0.81	0.22
15	0.53(0.07)	0.99	1.01	0.99	0.10
16	-0.28(0.06)	1.00	0.99	0.62	0.07
17	-0.16(0.07)	0.99	1.02	1.19	0.05
18	-0.42(0.08)	0.99	0.97	1.63	0.13
19	0.86(0.08)	0.98	1.02	1.23	0.20
20	0.43(0.07)	0.99	1.00	1.04	0.00
21	-0.17(0.06)	1.00	1.00	0.52	0.22
22	1.65(0.09)	0.99	1.05	0.94	0.01
23	-0.31(0.08)	1.00	0.99	1.14	0.02
24	-0.52(0.08)	0.99	1.05	1.40	0.03
25	-0.65(0.08)	0.99	1.12	1.41	0.03
26	-0.47(0.07)	0.99	1.05	1.04	0.02
27	-0.65(0.07)	1.00	1.00	0.96	0.28
28	-0.59(0.07)	0.99	1.01	1.10	0.08
29	0.95(0.07)	0.99	1.02	0.72	0.20
30	-0.44(0.08)	0.98	1.09	1.49	0.04
31	0.13(0.07)	0.99	1.01	1.07	0.23
32	1.01(0.07)	0.99	1.02	0.75	0.21
33	0.38(0.06)	1.00	1.01	0.64	0.21
34	0.72(0.08)	0.98	0.99	0.85	0.19
35	-0.40(0.07)	0.99	1.04	1.43	0.29
36	-1.35(0.1)	1.01	0.97	1.43	0.26
37	-0.80(0.07)	1.00	0.99	0.63	0.26
38	0.32(0.08)	0.99	0.99	1.36	0.14
39	0.55(0.08)	0.99	1.05	1.31	0.54**
40	0.53(0.09)	0.99	1.03	1.80	0.14
41	0.62(0.07)	0.99	1.01	0.96	0.12
42	-1.49(0.09)	1.01	1.01	1.00	0.20
43	-0.64(0.07)	1.01	1.00	0.71	0.22
44	0.95(0.07)	0.99	1.00	0.61	0.06
45	-0.77(0.08)	1.01	0.97	0.89	0.05
46	0.19(0.07)	0.99	1.00	1.03	0.22
47	0.68(0.08)	0.98	1.02	0.96	0.18
48	0.28(0.07)	0.99	1.01	0.92	0.13
49	0.47(0.07)	1.00	1.00	0.47	0.07
50	0.83(0.07)	0.99	1.01	0.47	0.12
51	1.28(0.07)	0.99	1.01	0.48	0.11
52	1.39(0.08)	0.99	1.02	0.55	0.24
53	0.23(0.08)	0.99	0.98	1.63	0.11
54	0.23(0.07)	1.00	0.98	0.82	0.32
55	1.36(0.07)	0.99	1.01	0.34	0.20
56	0.92(0.07)	1.00	1.00	-0.33*	0.71**
57	-0.07(0.06)	1.00	1.00	-0.20*	0.21
58	0.35(0.07)	1.00	1.00	0.81	0.32
59	2.12(0.1)	1.00	1.00	0.50	0.00
60	2.07(0.09)	1.00	1.00	0.22	0.11
61	0.30(0.06)	1.00	1.00	0.12	0.70**
62	1.19(0.07)	1.00	1.00	0.18	0.47
63	-0.34(0.07)	1.00	1.00	1.01	0.22
64	2.02(0.09)	1.00	1.00	-0.07*	0.07
65	1.73(0.08)	1.00	1.00	0.09	0.07
66	1.00(0.07)	1.00	1.00	0.54	0.48
67	-0.03(0.06)	1.00	1.00	0.12	0.06
68	-0.15(0.06)	1.00	1.00	0.48	0.13
69	0.88(0.07)	1.00	1.01	0.41	0.32
70	1.18(0.07)	1.00	1.00	0.42	0.16
71	1.51(0.10)	1.00	1.04	1.17	0.17
72	0.49(0.06)	1.00	1.01	0.44	0.03
73	1.36(0.07)	1.00	1.00	0.09	0.19
74	1.42(0.07)	1.00	1.00	-0.07*	0.06
75	0.09(0.05)	1.00	1.00	-0.04*	0.08
76	1.40(0.07)	1.00	1.00	-0.16*	0.01
77	1.67(0.08)	1.00	1.00	-0.15*	0.22
78	0.09(0.06)	1.00	1.00	0.49	0.29
79	1.77(0.08)	1.00	1.00	0.08	0.14
80	-0.51(0.06)	1.00	1.00	0.51	0.02

Note. *SE* = Standard error; DIF = Differential Item Functioning; * = poor discrimination value; ** = substantial DIF.



```
                  Dimension
----------------------------------------------
          Dimension_1         Dimension_2                      +item
---------------------------------------------------------------------------------------
                          |                   |                                       |
                          |                   |                                       |
                          |                   |                                       |
                         X|                   |                                       |
                         X|                   |                                       |
                          |                   |                                       |
                         X|                   |                                       |
   3                      |                 XX|                                       |
                          |                   |                                       |
                         X|                   |                                       |
                          |                   |                                       |
                         X|                   |                                       |
                        XX|                  X|                                       |
                        XX|                  X|                                       |
   2                     X|                  X|59 60 64                               |
                        XX|                 XX|                                       |
                       XXX|                 XX|79                                     |
                      XXXX|              XXXXX|22 65 77                               |
                      XXXX|              XXXXX|71                                     |
                     XXXXX|              XXXXX|52 55 73 74 76                         |
                    XXXXXX|            XXXXXXX|51                                     |
                  XXXXXXXX|            XXXXXXX|13 62 70                               |
   1                 XXXXX|           XXXXXXXX|12 29 32 44 66                         |
                 XXXXXXXXX|         XXXXXXXXXX|19 50 56 69                            |
                  XXXXXXXX|        XXXXXXXXXXX|4 34 47                                |
                XXXXXXXXXX|     XXXXXXXXXXXXXX|10 39 40 41                            |
                 XXXXXXXXX|     XXXXXXXXXXXXXX|3 15 20 49 72                          |
           XXXXXXXXXXXXXXX|    XXXXXXXXXXXXXXX|2 11 33 38 48 58 61                    |
            XXXXXXXXXXXXXX|     XXXXXXXXXXXXXX|1 46 53                                |
   0         XXXXXXXXXXXXX|     XXXXXXXXXXXXXX|31 75 78                               |
           XXXXXXXXXXXXXXX|    XXXXXXXXXXXXXXX|6 57 67                                |
          XXXXXXXXXXXXXXXX|    XXXXXXXXXXXXXXX|5 17 21 68                             |
             XXXXXXXXXXXXX|   XXXXXXXXXXXXXXXX|8 9 16 23 63                           |
      XXXXXXXXXXXXXXXXXXXX|        XXXXXXXXXXX|18 24 26 30 35 80                      |
           XXXXXXXXXXXXXXX|      XXXXXXXXXXXXX|25 27 28 43 54                         |
             XXXXXXXXXXXXX|         XXXXXXXXXX|45                                     |
            XXXXXXXXXXXXXX|       XXXXXXXXXXXX|37                                     |
  -1        XXXXXXXXXXXXXX|         XXXXXXXXXX|                                       |
              XXXXXXXXXXXX|            XXXXXXX|14                                     |
                  XXXXXXXX|            XXXXXXX|                                       |
                      XXXX|             XXXXXX|36                                     |
                      XXXX|              XXXXX|42                                     |
                        XX|                XXX|7                                      |
                        XX|                XXX|                                       |
  -2                     X|                XXX|                                       |
                          |                  X|                                       |
                          |                 XX|                                       |
                         X|                  X|                                       |
                          |                   |                                       |
                          |                  X|                                       |
                          |                  X|                                       |
                          |                   |                                       |
  -3                      |                   |                                       |
                          |                   |                                       |
                          |                   |                                       |
                          |                   |                                       |
                          |                   |                                       |
                          |                   |                                       |
                          |                   |                                       |
  -4                      |                   |                                       |
=======================================================================================
```

Figure 1 Map of latent distributions and responde model parameters.

Note. Each 'X' represents 4.4 cases

higher education systems, academic quality seems to be increasing. Hence, new public policies for student recruitment and selection should be able to obtain stronger support than in past years.

The application of common standard criteria in examination tests considers the variability of social background differentials in the enrolment and retention probabilities across student profiles (Jordan et al., 1996). This possibility allows improved actions in the environment based on a depth analysis of learning capitals (Ziegler & Baker, 2013). The main advantage is to establish more objective decisions regarding academic and professional trajectories that does not depend on the possible indirect cost of education or on less prestigious choices due to being more risk averse. In terms of episodic learning capital, resource investments imply that better-quality education and better involvement fosters cognitive and non-cognitive skills (Carneiro & Heckman, 2002; Ziegler, 2005). Latin-American institutions might decide on how to collaborate with national development using divergent strategies and consolidate an open policy without being worried about disadvantages, such us losing the effectiveness of the application of curriculum and knowledge through standard measurement. To this end, active purposes are intended to analyse the pragmatic and contextual approach to examining the process, the results obtained, and the methods used by various organisations (Sondergeld & Koskey, 2011) to ensure the principles of equity and equal opportunity for university admissions.

In conclusion, this study initiates an effective analysis in Ecuador that analyses test scores using advanced psychometric methods such as MIRT as an extension of European studies and American studies on official certificate examinations. However, it is important to bear in mind certain limitations, which may guide future research on this topic. First, it should be noted that the data used herein were students enrolled only in ESPN. Larger samples from other Ecuadorian universities may enable both better estimates of the achievement measures of students and deeper comparisons between the rates of access to levelling courses between the institutions. In this specific context, the use of MIRT models enable comparisons to determine appropriate measures for equity. Second, future analyses should include possible influences of the individual selection of science subjects or the effects of educational reforms on testing (Hübner et al., 2019; Korobko et al., 2008).

5 References

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis*, 23(2), 171-191. https://doi.org/10.3102/01623737023002171

Arocena, R. & Sutz, J. (2001). Changing knowledge production and Latin American universities. *Research Policy, 30*(8), 1221-1234. https://doi.org/10.1016/S0048-7333(00)00143-8

Arocena, R., & Sutz, J. (2005). Latin American Universities: From an original revolution to an uncertain transition. *Higher Education, 50*, 573-592. https://doi.org/10.1007/s10734-004-6367-8

Baird, J., Cresswell, M., & Newton, P. (2000). Would the real gold standard please step forward? *Research Papers in Education, 15*(2), 213-229. https://doi.org/10.1080/026715200402506

Bordieu, P., & Passeron, J.C. (1990). *Reproduction in education, society, and culture*. Sage.

Braxton, J., Sullivan, A., & Johnson, R. (1997). Appraising Tinto's theory of college student departure. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (pp. 107-164). Springer.

Breen, R., & Glothorpe, J. (1997). Explaining educational differentials: towards a formal rational action theory. *Rationality and Society, 9*, 275-305.

Carneiro, P., & Heckman, J. (2002). The evidence on credit constraints in postsecondary schooling. *Economic Journal, 112*, 705-734.

Carnoy, M., Khavenson, T., & Ivanova, A. (2015). Using TIMSS and PISA results to inform educational policy: a study of Russia and its neighbours. *Compare, 45*(2), 248-271.

Christensen, K. S., Oernboel, E., Nielsen, M. G., & Bech, P. (2019). Diagnosing depression in primary care: A Rasch analysis of the major depression inventory. *Scandinavian Journal of Primary Health Care,* 37, 105-112. https://doi.org/10.1080/02813432.2019.1608039

Coe, R. (2007). Common examinee methods for monitoring the comparability of examination standards. In P. Newton, J. Baird, H. Goldstein, H. Patrick and P. Tymms (Eds.), *Techniques for monitoring the comparability of examination standards.* London, UK: Qualifications and Curriculum Authority.

Coe, R. (2010). Understanding comparability of examination standards. *Research Papers in Education, 25*(3), 271-284. https://doi.org/10.1080/02671522.2010.498143

Contreras, F. A., & Maluk, S. A. (2017). Análisis descriptivo del gobierno universitario ecuatoriano: una mirada desde los cambios legislativos [descriptive análisis of the Ecuadorian university government: an análisis from the legislative changes]. *Revista Electrónica de Investigación educativa, 19*(2), 22-37. https://doi.org/10.24320/redie.2017.19.2.866

De Ayala, R. J. (2009). *The theory and practice of Item Response Theory*. The Guilford Press.

de Moura Castro, C., & Levy, D. C. (2001). Putting reality ahead of myths: A key to reform in Latin America. *International Higher Education, 22*, 16-18.

Di Caudo, M. V. (2015). Política de cuotas en Ecuador: me gané una beca para estudiar en la universidad. *Ponto-e-Vírgula, 17*, 196-218.

Embretson, S. E. (1984). *A general latent trait model for responses processes.* Psychometrika, 49, 175-186.

Eurydice (2012). *Key data on education in Europe 2012.* European Commission.

Fan, X. (1998). Item Response Theory and Classical Test theory: an Empirical Comparison of their Item/person statistics. *Educational and Psychological Measurement, 58*(3), 357-381. https://doi.org/10.1177/0013164498058003001

Georges, A., & Pallas, A. M. (2010). New look at a persistent problem: Inequality, Mathematics achievement, and teaching. The Journal of Educational Research, 103(4), 274-290. https://doi.org/1080/00220670903382996

Haladyna, T. M. & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice, 23*(1), 17-27. https://doi.org/10.1111/j.1745-3992.2004.tb00149.x

Hartig, J., & Höhler, J. (2009). Multidimensional IRT models for the assessment of competencies. *Studies in Educational Evaluation, 35*(2-3), 57-63. https://doi.org/10.1016/j.stueduc.2009.10.002

He, Q., Stockford, I., & Meadows, M. (2018). Inter-subject comparability of examination standards in GCSE and GCE in England. *Oxford Review of Education, 44*(4), 494-513. https://doi.org/10.1080/03054985.2018.1430562

Himmel, E. (2002). Modelos de análisis de la deserción estudiantil en la educación superior [models of analysis of student's dropout in Higher Education]. *Calidad de la Educación, 17*, 91-108.

Ho, E. S. C. (2016). The use of large-scale assessment (PISA): insight for policy and practice in the case of Hong Kong. *Research Papers in Education, 31*(5), 516-528. https://doi.org/10.1080/02671522.2016.1225351

Holland, P. W., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Hillsdale, NJ: Erlbaum.

Hübner, N., Wagner, W., Hochweber, J., Neumann, M., & Nagengast, B. (2019). Comparing apples and oranges: Curricular intensification reforms can change the meaning of students' grades! *Journal of Educational Psychology.* https://doi.org/10.1037/edu0000351

Jordan, W. L., Lara, J., & Mc Partland, J. M. (1996). Exploring the causes of early dropout among race-ethnic and gender groups. *Youth and Society, 28*(1), 62-94.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1-73. https://doi.org/10.1111/jedm.12000

Kaspersen, E., & Ytterhaug, B. O. (2020). Measuring mathematical identity in lower secondary school. *International Journal of Educational Research, 103,* 101620. https://doi.org/10.1016/j.ijer.2020.101620

Khorramdel, L., & von Davier, M. (2016). Item Response Theory as a Framework for test construction. in K. Schweizer & C. DiStefano (Eds.). *Principles and methods of test constructions* (pp. 52-82). Göttingen: Hogrefe.

Korobko, O. B., Glas, C. A., Boskeer, R. J., & Luyten, J. W. (2008). Comparing the difficulty of examination subjects with item response theory. *Journal of Educational Measurement, 45*(2), 139-157. https://doi.org/10.1111/j.1745-3984.2007.00057.x

Lietz, P., & Tobin, M. (2016). The impact of large-scale assessment in education on education policy: evidence from around the world. *Research Papers in Education, 31*(5), 499-501. https://doi.org/10.1080/0267122.2016.1225918

Liou, P. Y., & Bulut, O. (2020). The effects of item format and cognitive domain on students' science performance in TIMSS 2011. *Research in Science Education, 50*, 99-121. https://doi.org/10.1007/s11165-017-9682-7

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Lawrence Erlbaum.

Ma, Y., Hou, X., Huang, J., Wang, W. Li, Y., Zhou, X., & Du, X. (2018). Educational inequality and achievement disparity: An empirical study of migrant children in China. *Children and Youth Services Review, 87,* 145-153. https://doi.org/10.1016/j.childyouth.2018.02.026

Reckase, M. D. (2009). *Multidimensional Item Response Theory*. Springer.

Ribeiro, D. (1971). *La Universidad Latinoamericana [The Latin-American University]*. Caracas: Ediciones de la Biblioteca de la Universidad Central de Venezuela.

SENESCYT (2018). *Proceso de Admisión 2018 [Admission Process 2018]*. Available at: https://admision.senescyt.gob.ec/soluciones/existe-puntaje-minimo-ingresas-una-carrera/.

Schmidt, W. H., & Burroughs, N. A. (2016). Influencing public school policy in the United States: the role of large-scale assessments. *Research Papers in Education, 31*(5), 567-577. https://doi.org/10.1080/02671522.2016.1225355

Sondergeld, T., & Koskey, K. (2011). Evaluating the impact of an urban comprehensive school reform: An illustration of the need for mixed methods. *Studies in Educational Evaluation, 37*, 91-107. https://doi.org/10.1016/j.stueduc.2011.08.001

Tünnermann, C. (2000). *Universidad y Sociedad. Balance histórico y perspectivas desde Latinoamérica [University and society. Historical balance and perspectives from Latin-America]*. Caracas: Universidad Central de Venezuela.

Van der Linden, W. J. (2017). *Handbook of item response theory, volume three: Applications*. Chapman and Hall/CRC. https://doi.org/10.1201/9781315117430

Veas, A., Benítez, I., Navas, L., & Gilar-Corbí, R. (2020a). A comparative analysis of university entrance examinations using the construct comparability approach. *Revista de Educación, 388*, 65-83. https://doi.org/10.4438/1988-592X-RE-2020-388-447

Veas, A., Navas, L., Pozo-Rico, T., & Miñano, P. (2020b). University Entrance Examinations in Spain: using the construct comparability approach to analyze standards quality. *Frontiers in Psychology, 11*: 127. https://doi.org/10.3389/fpsyg.2020.00127

Wang, W. C., Yao, G., Tsai, Y. J., Wang, J. D., & Hsich, L. (2006). Validating, improving reliability, and estimating correlation of the four subscales in the WHOQOL-BREF using multidimensional Rasch analysis. *Quality of Life Research, 15*, 607-620. https://doi.org/10.1007/s11136-005-4365-7

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). *ACER Conquest, version 2.0: Generalized item response modelling software.* Camberwell: Australian Council for Educational Research.

Ziegler, A., & Baker, J. (2013). Talent development as adaption: The rôle of educational and learning capital. In S. Phillipson, H. Stoeger, & A. Ziegler (Eds.), *Exceptionality in East-Asia: Explorations in the actiotope model of giftedness* (pp.18-39). Routledge.

Ziegler, A., Chandler, K., Vialle, W., & Stoeger, H. (2017). Exogenous and endogenous learning resources in the Actiotope Model of Giftedness and its significance for gifted education. *Journal for the Education of the Gifted, 40*, 310-333. https://doi.org/10.1177/0162353217734376

Ziegler, A., Debatin, T., & Stoeger, H. (2019). Learning resources and talent development from a systemic point of view. *Annals of the New York Academy of Sciences*, 1445, 39—51. https://doi.org/10.1111/nyas.14018

Ziegler, A., & Stoeger, H. (2017). Systemic gifted education. A theoretical introduction. *Gifted Child Quarterly, 61*, 183–193. doi:10.1177/0016986217705 713

Ziegler, A., Stoeger, H., & Balestrini, D. (2017). Systemic gifted education. In J. Riedl Cross, C. O'Rellly & T. Cross (Eds.), *Providing for the special needs of students with gifts and talents* (pp. 15-55). Dublin, Ireland: Kazoo Independent Publishing.

Ziegler, A., Vialle, W., & Wimmer, B. (2013). The actiotope model of giftedness: A short introduction to some central theoretical assumptions. In S. Phillipson, H. Stoeger, & A. Ziegler (Eds.), *Exceptionality in East Asia: Explorations in the actiotope model of giftedness* (pp. 1-17). London, England: Routledge. Ziegler, A., Vialle, W., & Wimmer, B. (2013). The actiotope model of giftedness: A short introduction to some central theoretical assumptions. In S. Phillipson, H. Stoeger, & A. Ziegler (Eds.), *Exceptionality in East Asia: Explorations in the actiotope model of giftedness* (pp. 1-17). London, England: Routledge.

Declarations

Funding
Not applicable.

Conflicts of interests
The authors declare that there is not any conflict of interest.

Authors' contribution
Iván Sandoval: Theoretical review, data collection, manuscript review.
Raquel Gilar-Corbí: Theoretical Review, data collection.
Alejandro Veas: Theoretical review, data analyses.
Juan Luis Castejón: manuscript review.

Corresponding author:
Alejandro Veas, PhD
University of Alicante
Department of Developmental Psychology and Didactics
PO 03690, San Vicente del Raspeig
Alicante, Spain
E-mail: Alejandro.veas@ua.es