REPeC, Brasília, v. 10, n. 3, art. 6, p. 323-345, Jul./Sep. 2016
Available online at *www.repec.org.br*
DOI: http://dx.doi.org/10.17524/repec.v10i3.1401

Revista de Educação e Pesquisa em Contabilidade
Journal of Education and Research in Accounting

Periódico Trimestral, digital e gratuito publicado pela Academia Brasileira de Ciências Contábeis

ISSN 1981-8610

Performance of Accounting students on the Enade/2012 test: an application of the Item-Response Theory

Abstract

The objective in this study was to measure Accounting students' performance (proficiency) on the Enade test using the Item Response Theory (IRT). The students' performance was measured using the three parameter logistic model (3PL), based on data related to the Enade test/2012, taken from the website of the National Institute for Educational Studies and Research Anísio Teixeira (Inep), concerning 47,098 students. Through the scale, three levels of student performance could be distinguished. Level 1 students master the reading and interpretation of texts and quantitative reasoning. In addition, Level 2 students should present logical reasoning and systemic and holistic perspective. Furthermore, at Level 3, students should present interdisciplinary knowledge, covering accounting contents, critical-analytic skills and practical application of the content mastered. The results also appointed that the items of the Enade test were very difficulty for the group that took the test. Independently of the student characteristics analyzed, overall, the proficiency scores were very low. This result suggests that the HEI need to take actions and that public policies are needed that can contribute to improve the students' performance.

Key words: Student performance; Proficiency; Enade; Item Response Theory.

Raphael Vinicius Weigert Camargo
M.Sc. and Ph.D. candidate in Accounting at Universidade Federal de Santa Catarina (UFSC) and Professor at Universidade Estadual do Paraná (Unespar). **Contact:** Rua Comendador Correa Junior , 117. Centro. Paranaguá-PR. CEP.: 83203-560.
E-mail: raphael.camargo@unespar.edu.br

Rita de Cássia Correa Pepinelli Camargo
M.Sc. and Ph.D. candidate in Accounting at Universidade Federal de Santa Catarina (UFSC). **Contact:** Campus Universitário Trindade, Departamento de Ciências Contábeis, Bloco C – sala 215. Bairro: Trindade. Florianópolis-SC. CEP.: 88040-970.
E-mail: rita.camargo@ufsc.br

Dalton Francisco de Andrade
Ph.D. in Biostatistics from University of North Carolina at Chapel Hill, NC – USA and Full Professor at Universidade Federal de Santa Catarina (UFSC). **Contact:** Campus Universitário Trindade, Centro Tecnológico, Departamento de Informática e Estatística, sala 509. Bairro: Trindade. Florianópolis-SC. CEP.: 88040-900.
E-mail: dalton.andrade@ufsc.br

Antonio Cezar Bornia
Ph.D. in Production Engineering from Universidade Federal de Santa Catarina (UFSC) and Full Professor at Universidade Federal de Santa Catarina (UFSC). **Contact:** Campus Universitário Trindade, Centro Tecnológico, Departamento de Engenharia de Produção e Sistemas, CP 476. Bairro Trindade. Florianópolis-SC. CEP.: 88040-900.
E-mail: cezar.bornia@ufsc.br

Published in Portuguese and English. Original Version in Portuguese.

Received in 3/23/2016. Ask to Revise on 5/27/2016. Resubmitted on 6/20/2016. Accepted on 6/22/2016 by Dr. Márcia Maria dos Santos Bortolocci Espejo (Assistant Editor) and by Dr. Valcemiro Nossa (Editor). Published on 31/7/2016. Organization responsible for the journal: Abracicon.

1. Introduction

Education can be understood as a scarce economic good, which depends on choices and renunciations of individual or collective resources (Waltenberg, 2006). According to Cunha, Cornachione Junior and Martins (2010), empirical studies show that the importance of education and its effects are higher in developing countries compared with developed countries. The establishment and maintenance of public policies related to financial education systems, educational accountability and educational integration depend mainly on monitoring the situation the education is offered in, usually performed through assessments (Hanushek, 1979).

According to Hanushek (1979), educational knowledge assessment tests are used as a means to measure the output of the educational production process. For Primi, Hutz and Silva (2011, p. 271), "in democratic societies, such assessments serve as accountability tools, used to determine whether the organizations, to whom some roles are attributed, have fulfilled their obligations".

As a scarce economic good, a parameter for measuring its current stage is the quality of education, a parameter that supports the evaluation of the effectiveness of teaching. In Brazil, higher education is evaluated periodically by education monitoring entities subordinated to the Ministry of Education (MEC). In the case of undergraduate education, this monitoring is carried out by the National System of Higher Education Assessment (Sinaes), which adopts the National Examination of Student Performance (Enade), established in 2004 as a tool for evaluating student performance in higher education courses in relation to the program contents and skills (Inep, 2014). In addition, "the guidelines for the preparation of the Accounting test are defined in Inep Ordinance 202 of June 22, 2012" (Inep, 2014, p. 6). It should be noted that, in addition to the objective set by Sinaes to evaluate the students' performance, the Enade test is also adopted to evaluate the performance of educational institutions in imparting knowledge for students' education in relation to compliance with the curriculum components of courses based on a National Curriculum Guideline.

This test has been used, in addition to the purposes set by Sinaes, as a measure of higher education students' performance in scientific research. These surveys, usually carried out within specific areas of expertise, such as Psychology, Pedagogy, Administration, among other areas, investigate the situation of teaching in those respective areas. These evaluations, based on performance measures, are elaborated with a view to actions to improve education and the understanding of (determining) factors related to higher (or lower) academic performance, measured based on the tests.

In the scientific literature, which shows measures of higher education students' academic performance, one can see the predominance of studies using the classical test theory (CTT) to measure the students' knowledge. Based on these measures, the needs to improve the education offered are evaluated, suggestions are elaborated to improve the teaching and the factors are understood the factors that help to explain the reasons why the performance tends to be higher or lower. However, through the classical analysis, this survey may not express the reality of academic performance, as it does not consider the specifics of each item in the evaluation.

In the fields of psychology and pedagogy, respectively, studies by Primi et al. (2011) and Lopes and Vendramini (2015) used the Item Response Theory (IRT) as an alternative to measure students' performance (proficiency) on the Enade tests. It is observed in these studies that the assessment based on IRT presents more detailed results on the proficiency of the individuals assessed, also offering information about which knowledge they specifically mastered, in which both the item and the respondent are placed on the same measuring scale. Thus, among other possible observations, one can mention the comparison of the difficulty of proof regarding the knowledge (proficiency) of the student group.

Unlike TCT, IRT permits measuring students' knowledge (proficiency) according to the complexity of the content that makes up each exam question. Based on this measure, the proficiency the student effectively demonstrated can be identified, since it not only considers the number of items answered correctly. From this perspective, it is possible to distinguish between individuals, so that different students who have correctly answered the same number of questions (items), with different degrees of difficulty, have different proficiency measures.

This study seeks to contribute to Accounting teaching in Brazil by presenting an alternative measure of students' performance in the area through IRT. Thus, this study aims to measure the Accounting students' performance (proficiency) on the Enade by means of IRT.

Through this study, the intent is to model the probability of correctly answering the questions (items) that make up the test, considering the student's proficiency based on the difficulty and distinctive capacity of the items and the likelihood of a casual hit. As it permits assessing the knowledge the student effectively masters in each item, this initiative enables interventions and timely corrective actions in the teaching-learning process in order to identify gaps that comprise possible learning disabilities.

Considering education as a scarce resource, its importance in developing countries and the search to develop public policies related to education, the problem of this study rests on the Economic Theory of Education. This study aims to assess the situation of higher education in the field of Accounting through the students' performance, which corresponds to one of the end products of education, as graduates. It is necessary to evaluate and understand if the "education" resource is managed to provide private and social benefits.

Thus, the intention is to contribute to the field of education in Accounting by seeking alternative tools to measure the current situation of higher education in Accounting in Brazil, based on the evaluation of students' academic performance (proficiency), in order to assess the situation found in each type of knowledge that integrates the curriculum components of higher education courses in Accounting in detail.

The article is divided into five sections. The next section presents the theoretical framework, which includes the Economic Theory of Education, item response theory and previous studies related to the topic of the article; the third section presents the methodological procedures used in the research; the fourth section presents the results of the study, considering the estimation of the parameters of the model adopted, to measure the proficiency of students and the proficiency scale built. In the last section, the conclusions are exposed and suggestions are provided for future studies.

2. Theoretical Framework

2.1 Economic Theory of Education – Human Capital

Blaug (1992) classifies research based on Economic Theory of Education in two major fields. The first is related to the analysis of the economic value of education and the second investigates the economic aspects of education systems. Studies in the first field have concluded that the economic growth of a country or region is closely linked to people's level of education (Waltenberg, 2006). The theoretical aspects of human capital are part of this first field of the Economic Theory of Education.

The Human Capital Theory essentially seeks to understand and explain the consequences of education for the value of people's human capital. In this current, the gain of knowledge and skills causes an increase in the value of individuals' human capital, reflected in employment, productivity and potential income (Cunha et al., 2010). From this perspective, education can be considered an economic good, with problems of scarcity and dependent on choices and sacrifices in terms of individual and collective resources (Waltenberg, 2006).

This idea derives from the views of Schultz (1960), considered one of the forerunners of the Human Capital Theory. For the author, education can be treated as a form of investment in human beings and will add value to the individual who receives it. In this conception, yield and productivity increases are the result of increased human capital. Human capital, while investment in the human being is composed of two types of resources. The first relates to the resources invested by educational institutions, and the second to investments made by individuals in the form of gains waived as students. This last feature has been observed empirically by Schultz (1960) as the most significant part of the costs incurred in student education.

Later, Becker (1962) has contributed to the expansion of the Human Capital Theory, through empirical studies of the consequences of that capital gain on issues like welfare and wealth. According to Waltenberg (2006), the results found in studies conducted in this area still lack a direct quantitative relationship between economic growth and education. The definition of an optimal level of education in a country does not exist. There is no consensus on a minimum level of education that makes a country reach a condition of rapid growth, nor are the maturity and depreciation periods of human capital known. Rasera (1999) points out that education has a problem related to its form of measuring, which makes it difficult to measure a reasonable relationship between economic growth and investment in human capital. Thus, one of the few possible conclusions is that education fosters the growth of a country, but without the degree of relationship between these two variables.

The measuring and understanding of the determinants of educational performance are important for the development of policies such as educational accountability, financial education systems and educational integration. Most studies in this area used standardized outputs obtained from test marks as a performance measure (Hanushek, 1979). What is perceived in such studies, however, is the predominance of educational performance measures based on the classical test theory (CTT). This form of measuring has some weaknesses, especially with regard to performance evaluation depending on the particularities of each item in the assessment tool, specifically the tests carried out by students.

Studies like Primi et al. (2011) and Lopes and Vendramini (2015) employed the Item Response Theory (IRT) as an alternative to measure students' performance (proficiency) on Enade tests. Using the Item Response Theory (IRT), presented in further detail in subsequent sections of this study, a measure can be obtained to assess the students' performance, measuring and distinguishing the knowledge the student presented in the test. In this sense, the IRT can contribute to the human capital current in the development of the Economic Theory of Education.

These studies contribute by providing feedback on the quality of education offered to the population. This is because there is a need for greater effectiveness in teaching, mainly due to its effects in different sectors and strata of the socio-political and economic organization of a country (Lopes & Vendramini, 2015).

2.2 Item Response Theory (IRT)

The IRT includes a set of modeling techniques that, through the analysis of specific items, permit measuring certain characteristics of the individuals (Edelen & Reeve, 2007). Through these techniques, in accordance with Andrade, Tavares and Valle (2000), the relationship between an individual's latent trait(s) can be represented and the likelihood that (s)he will answer this particular item correctly. Latent trait is considered to be the characteristics of an individual that cannot be directly observed, such as proficiency in English or consumer satisfaction. An item can be represented by a matter of proof, or even by a question related to the satisfaction of the individual with a particular product or service. In the case studied in this article, latent trait refers to performance (proficiency), expressing accounting knowledge, measured by the number of items (questions) in the Enade test / 2012.

In recent decades, there has been an increase in the application of techniques derived from IRT, mainly in educational assessment (Andrade et al., 2000). The first works related to this theory emerged in the 1950's, with Lord (1952). Next came the work of Rasch (1960), Birnbaum (1968), Wright (1968) and Samejima (1969), the latter with the proposal of a gradual response model, as previous models were only applied to dichotomous responses. Nowadays, IRT models are widely applied around the world, and more information about the origin of this theory can be found in the works of Andrade et al. (2000), Bock (1997), Boomsma, Van Duijn and Snijders (2000), Van Der Linden and Hambleton (1997) and Moreira (2010).

The resulting IRT models are able to meet some limitations of the Classical Test Theory (CTT) or Classical Measurement Theory (CMT), which is commonly used in the evaluation and selection of individuals. In CTT, evaluation is based on raw and standardized scores. The score of an individual depends

primarily on the number of hits achieved, regardless of whether these correct answers are related to easy or difficult questions (Andrade et al., 2000).

One of the differences that can be identified between the CTT and the IRT is in how the models deriving from IRT treat the test. Taking the field of educational assessment as an example, in IRT, the evaluation of the test items, i.e. the test is considered on the whole, while items are considered individually in CTT, so that the score achieved on an item is independent of the score achieved on other items. The benefits IRT provides in this case is that it can capture subjects who respond to questions more consistently, that is, hitting easier instead of difficult questions, which values their scores. On the other hand, it punishes the score of individuals who hit more difficult than easy questions, since the logic is that those who get the most difficult questions right should also master the easier questions, which may be related to the random hit (Moreira, 2010).

Andrade et al. (2000) emphasize that the greatest advances of IRT are related to the creation of interpretable scales; with the possibility of comparison between individuals who have made different tests, but with items belonging to the same scale; and the comparison between individuals of different populations, who conducted tests with some items in common.

In Brazil, in addition to the wide application of IRT models in the educational assessment area, the theory has been used in several other areas, such as the degree of consumer satisfaction, valuation of companies' intangible assets, organizational management, evaluation of quality of life, psychological assessments, among others, which can be found in Moreira's work (2010). According to the author, studies have used various one-dimensional IRT models, especially the three parameter logistic model (3PL). In this study, the 3PL is used. This model considers item discrimination, item difficulty and probability of casual hit parameters, unlike the Rasch model, which only takes into account the difficulty. In the next section, some examples are discussed of studies that used the IRT to analyze the Enade test, as well as Accounting studies related to Enade.

2.3 Empirical studies

Internationally, Bartlett, Peel and Pendlebury (1993) found that the performance of students in the third year of Accounting at the University of Wales College of Cardiff was more related to the performance in the first year than other characteristics, such as demographic, financial or investments features. Byrne and Flood (2008) observed the overall performance increase in Financial and Management Accounting among students from Dublin City University with background knowledge in accounting. Both studies used the classical theory for performance measurement, based on the average students achieved in the subjects Financial and Management Accounting.

Davidson (2003), in an analysis of Canadian students, noted that superior performance in Accounting on exams that include complex issues is related to a more in-depth study approach and not just superficial memorization techniques. Factors such as previous academic performance, motivation to accomplish the course and intention to work in accounting also help to explain performance on assessment tests. Unlike the present study, student performance was measured based on an application of Classical theory (classical analysis), weighted by the relative level of difficulty involved in each question. Although this is one way to minimize the distortions arising from the use of basic Classical theory, in which all questions have the same weight in the evaluation, this level of difficulty does not fit the specific characteristics of the group of students who performed the test.

Clinton and Kohlmeyer (2005) evaluated US students and found that, although there was no significant difference between the performance of individuals submitted to an educational process based on quizzes and processes based on conventional methods, there was a significant difference in the learning perception and motivation to study in the group of students undergoing the quizz-based method. Like Davidson (2003) and unlike the present study, they used the Classical theory to measure student performance. One caveat about Clinton and Kohlmeyer's approach (2005) is that the lack of significance in the

difference between the teaching approaches can derive from distortions due to the use of Classical theory. The possibility to reassess the effectiveness of quizz-based study is the measuring of the dependent variable (student performance) based on proficiency estimated by the IRT.

In Brazil, some studies used the Enade test as the base to measure the performance of Accounting students. Souza (2008) reviewed the performance determinants of students on the Enade test / 2006. He found that variables such as the students' background knowledge prior to his admission to the course, the father's education, personal effort and family income influence the performance of the evaluated students. Santos (2012) extended the study by Souza (2008) and analyzed the effect of individual and institutional variables on student performance and data from the National Survey of courses from 2002 and 2003 and Enade / 2006, noting that variables such as gender, hours devoted to study, family income and attendance of public high school interfere positively in student performance. In addition, institutions whose teaching staff held Master's and Ph.D. degrees and worked on a full-time contract also contribute to higher student achievement.

Miranda (2011) investigated the relationship between the performance of students and the training of teachers in Accounting courses. The author used the results of Enade / 2009 as the base to measure student performance results. The results show that the academic qualification of teachers at the Higher Education Institutions (HEIs) investigated can influence the performance of students on the Enade test. Similarly, Santos, Cunha and Cornachione Junior (2009) observed that the concept of Accounting courses partially reflects the teachers' degree. The authors used a nominal five-category variable to evaluate the course performance. It should be noted that prior to the conversion to nominal, the measuring form used comes from Classical theory. Cruz, Nossa, Balassiano and Teixeira (2013) found no relationship confirming that the didactic and pedagogical structure of curricula influences the performance of students on the Enade / 2009. It was only confirmed that the geographical region of the course influences the performance.

Although there are studies related to the ENADE test in the Brazilian accounting literature, no studies were identified that investigate the relationships between the test items and student proficiency (performance) based on the IRT. In the studies reviewed, only student performance measures based on Classical theory were found, which, as appointed in the previous section, presents a number of disadvantages compared to IRT.

Some studies in other knowledge areas analyzed the Enade with the application of item response theory (IRT). Examples include Management (Scher, Moreira, Correa, Schuch, Andrade & Bortolotti, 2014), Statistics (Coelho, Ribeiro Junior & Bonat, 2014), Psychology (Primi et al., 2011) and Education (Lopes & Vendramini, 2015).

Scher et al. (2014) analyzed the Enade / 2009 taken by Management students, using the 3PL. The results showed the feasibility of using the IRT for the Enade items evaluated and also noted the increased proficiency between freshmen and seniors. Coelho et al. (2014) applied the IRT to students who took the Enade / 2009 in Statistics and found that few students showed high proficiency on the test, concentrating most respondents at an average proficiency level. Primi et al. (2011) analyzed the Enade / 2006 taken by new students and graduates in Psychology, using factor analysis and Rasch' model; they presented the construction of four performance ranges, describing, based on the test items, what knowledge, skills and expertise were needed for the student to demonstrate proficiency in each range; and observed a concentration of graduating students in higher proficiency ranges and that, in courses with higher Enade scores, the difference in performance between new and graduating students is smaller. Lopes and Vendramini (2015) analyzed the Enade test / 2005 taken by Pedagogy students. Similarly to the study by Primi et al. (2011), they also used the Rasch model and found that the test includes questions of median difficulty, similar to students' average and therefore appropriate proficiency.

It is observed among studies in other areas that used the IRT that the assessment made contains more detailed ways of measuring students' proficiency, in which both the test items and the respondents are placed on the same measuring scale, so that the knowledge the students specifically master can be indicated, The application of IRT to the Enade test of Accounting students can also contribute to understand the knowledge and skills the investigated students master, according to their level of proficiency.

3. Methodological Procedures

To measure the accounting students' performance (proficiency) on Enade/2012, the 3PL of IRT was used. The analyses were developed based on the estimation by means of BILOG-MG® software and the graphs produced in R software (R Development Core Team, 2012).

3.1 Population and sample

The study population consists of senior students of Accounting courses in Brazil. In total, 57,248 students enrolled to take Enade/2012 (INEP, 2014). Initially, 47,124 students with valid data were identified, regularly enrolled in Accounting courses in Brazil, convened by the Ministry of Education (MEC) and who attended the examination. Students were selected the students who took all objective tests, both General and Specific Knowledge; and noted 26 students who attended but handed back the test unsolved, i.e. they did not answer the questions. Thus, the final sample consists of 47,098 students who took the Enade / 2012. Data were extracted from the website of the National Institute of Educational Studies and Research Anísio Teixeira (INEP) on October 26th, 2015.

The age of the students ranged from 19 to 79 years, with an average of 29 years. The largest share of students was between 21 and 25 years old (19,756 people - 42%); 12,556 (27%) of the students were between 26 and 30 years; 7,135 (15%) were between 31 and 35 years; and 7,534 (16%) over 35 years. Only 177 students were up to 20 years old. Most students (94%) studied at night. There was predominance of female students (59%) and from the Southeast (39%) and South (24%). The North region concentrated the lowest number of students (6%), followed by the Central-West (13%) and Northeast (18%). Of the students in the sample, 16% came from public and 84% from private Higher Education Institutions (HEIs).

3.2 Measuring of student performance (proficiency)

3.2.1 Enade test

The Enade test / 2012 consists of discursive and objective questions (both general knowledge and specific knowledge). On the whole, there are forty (40) questions, divided into ten (10) general education questions (two discursive and eight multiple choice) and thirty (30) specialized questions (three discursive and twenty-seven multiple choice).

The general education questions involve knowledge of art and culture, technological advances, democracy, ethics and citizenship, ecology and biodiversity policies, labor relations, social responsibility, among others. The purpose of this group of questions is to assess characteristics related to an ethical, competent professional who is committed to the society he lives in (INEP, 2014).

The specific knowledge component involves questions from Accounting Theory, Professional Ethics, Financial Accounting, Cost Accounting and Analysis, Management Accounting and Controllership, among others. The purpose of these questions is to assess the skills of students in using Accounting terminologies and language, present systemic and interdisciplinary perspectives, mastery of the identification, recognition, measurement and disclosure phases, critical and analytical skills, production of information for decision making, interpretation and application of accounting standards, among other competences (INEP, 2014).

For the study, exclusively the objective questions were considered, i.e. 35 (thirty five) questions (eight general education and twenty-seven specialized), which were treated as dichotomized items and evaluated using the Three Parameter Logistic Model of the IRT (3PL), presented in further detail in section 3.2.2.

3.2.2 IRT Measuring Model

The performance measure was developed by means of IRT, in which a scale was built to evaluate the students' proficiency levels in Enade / 2012. In this form of measuring, all items that make up the assessment tool (test) are placed on the same proficiency measuring scale. Through this step, all test items could be placed on "anchor" levels to interpret the required test performance.

According to Andrade et al. (2000), the models proposed in the literature, originating in IRT, are based on three factors, namely: (i) the nature of the item, which can be dichotomous or not dichotomous; (Ii) number of people involved, represented by one or more than one; and (iii) measured amount of latent features, which can be one or more (one-dimensional or multidimensional).

In this study, we used a one-dimensional model, represented by the latent trait "knowledge in accounting", a single population is considered, i.e. Accounting students who took the Enade test / 2012; and dichotomized items are used.

The 3PL is applicable when the items are dichotomous or polytomous (items with two or more categories) with a single correct answer option. In the latter case, the polytomous items should be adjusted (dichotomized) into two categories, that is, right and wrong. This model is applicable in case you want to evaluate the item difficulty, discrimination and the probability of individuals with low skills levels giving a correct answer (Andrade et al., 2000).

The 3PL is displayed in Figure 1:

$$P(U_{ij} = 1|\theta_j) = c_i + (1 - c_i)\frac{1}{1 + e^{-a_i(\theta_j - b_i)}}$$

with $i = 1, 2, 3, \ldots, I$, and $j = 1, 2, 3, \ldots, n$,

In which:

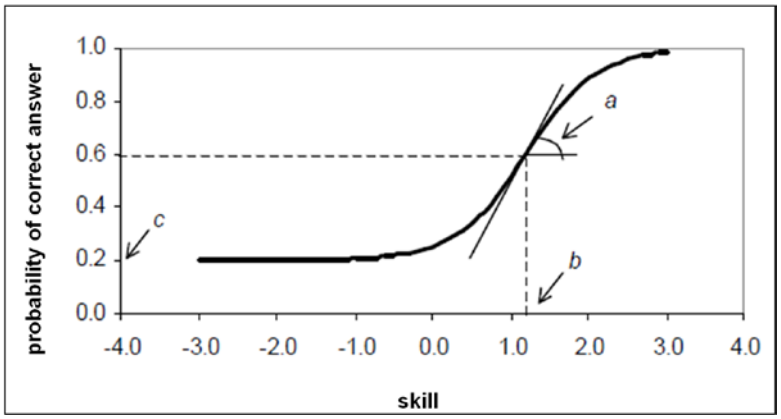
U_{ij}: dichotomous variable equal to 1 (one) when individual j gives a correct answer to item i, or 0 (zero) in the opposite case;
θ_j: Represents the skill (latent trait) of the j-eth individual investigated;
$P(U_{ij} = 1|\theta_j)$: Is the probability that individual j with skill θ_j will answer item i correctly and is called Item Response Function (IRF);
a_i: Represents the discrimination parameter of item i. The value of the parameter is proportional to the inclination of the Item Characteristic Curve – (ICC) in point b_i.
b_i: represents the difficulty parameter of item i, measured on the same scale of skills;
c_i: is the item parameter that represents the probability that individuals with low skills will correctly answer item i, that is, it represents the probability of a random hit.

Source: adapted from Andrade *et al.* (2000, pp. 9-10).

Figure 1. One-dimensional Three Parameter Logistic Model (3PL)

To estimate the model parameters based on IRT, generally, the Marginal Maximum Likelihood Method is applied, jointly applying an iterative process like the Newton-Raphson or Fisher Scoring algorithm, or Bayesian procedures, in accordance with Andrade *et al.* (2000).

In Figure 2, the Item Characteristic Curve (ICC) is displayed.



Source: Andrade *et al.* (2000, p. 11).

Figure 2. Item Characteristic Curve (ICC)

The ICC represents the existing relation between the probability of an individual, given his/her skill to correctly answer a certain item ($P(U_{ij}=1|\theta_j)$) and the parameters of the 3PL a_i, b_i and c_i, which represent, respectively, the inclination of the curve (discrimination), the position of the item in the scale (difficulty) and the possibility of a random hit for individuals with low skills. Axis Y represents the probability of an individual's correct answer, while axis X demonstrates the skill related to the latent trait assessed.

In IRT, according to Andrade *et al.* (2000), the individual proficiency (skill) measured can correspond to a real value between -∞ and +∞. In this case, an origin and a measuring unit need to be established for the scale that will be constructed for the latent trait assessed. The standard scale generally used in software that estimates the IRT models is represented by the mean 0 and standard deviation 1, that is, scale (0,1). This scale can be modified, provided that the existing relations between the model parameters are maintained. Information on the transformation of scales can be consulted in Andrade *et al.* (2000). In this study, the scale (0,1) was adopted to analyze the Accounting students' performance on Enade/2012.

In 3PL, the parameter a_i represents the discrimination of the item assessed. Items with excessively low coefficients on this parameter indicate a low discrimination power, that is, the probability that individuals with low proficiency will correctly answer the item is the same as for individuals with high proficiency. In general, according to the scale used (0,1), for an item to possess good discrimination power, the parameter a_i should be superior to 1 (Andrade *et al.*, 2000), although items with parameters superior to 0.7 are also acceptable (Scher *et al.*, 2014). That is so because the parameter a_i is proportional to the derivative of the tangent of the logistic function curve in the inflection point. Hence, steeper curves (with high coefficients for this parameter) basically distinguish individuals in two groups: individuals who master the knowledge of the item and individuals who do not. On the other hand, lower coefficients for this parameter indicate the item's low discrimination power, in which students with different skills have practically the same probability to correctly answer the item (Andrade *et al.*, 2000).

b_i is the parameter of the item's difficulty – the higher, the more difficult to item is and vice-versa. This means that, for individuals with a higher proficiency level, the probability of a correct answer is higher. It is also known as the location parameter, as it helps to verify the position of the item on the proficiency scale (Scher *et al.*, 2014). For scale (0,1), which considers an average of zero and standard deviation of one, the item coefficients for the parameter are expected to range between -2 and +2 (Andrade *et al.*, 2000).

The parameter c_i represents the probability that a low-skilled individual will correctly answer a given item. Its coefficient represents the number of alternatives in the item. Hence, for an item consisting of five alternatives, the theoretical probability that an individual will correctly mark any of the alternatives equals 0.2, a margin of tolerance can be established (superior and inferior to 0.2), in which the student's answer can be considered a mere casual hit, instead of the mastery of knowledge on the item. Thus, the expected coefficient for the parameter ranges between 0.1 and 0.3 (Scher *et al.*, 2014).

After measuring the students' proficiency and analyzing the parameters resulting from the estimated model, the next step was to analyze the students' performance in the research population. Therefore, in the next topic, the interpretation of the scale is discussed.

3.3 Analysis of student performance (proficiency)

Based on the measure of the students' performance, all questions were placed on the same scale, ranked according to the parameter of item difficulty (b_i). Items that simultaneously complied with the criteria displayed in Figure 3 were considered as anchor items.

$$
\begin{aligned}
&1.\ P(U = 1 \mid \theta = Z) > 0.65;\\
&2.\ P(U = 1 \mid \theta = Y) < 0.50;\ and\\
&3.\ P(U = 1 \mid \theta = Z) - P(U = 1 \mid \theta = Y) > 0.30.
\end{aligned}
$$

Source: Andrade *et al.* (2000, p. 110).

Figure 3. Criteria to define "anchor" levels

Thus, for an item to be considered an "anchor", most individuals (65%) need to answer it correctly, with the proficiency indicated by the level of the scale, and by a lesser proportion of individuals, i.e. maximum 50%, of individuals with a lower proficiency level. In addition, a 30% difference needs to exist between individuals with different skill levels who correctly answered the item (Andrade *et al.*, 2000). Due to the difficulty to simultaneously comply with the three criteria, however, items that complied with at least two of the criteria mentioned were also considered as "almost anchor" items.

The items identified as "anchor" or "almost anchor" items, previously defined according to Figure 3, were then analyzed and related to the objectives of the question, to the required competences and to the knowledge objects involved, as established by Inep in the Assessment Matrix of the Enade test / 2012 in Accounting, displayed in Table 1. This analysis was developed in that manner because, statistically, the IRT can distinguish among the proficiency levels and, pedagogically, the knowledge and competences the test requires in each proficiency level can be appointed.

Table 1

Assessment matrix of Enade test/2012 in Accounting

Classificação	Componente de Formação Geral	Componente Específico
Objetivos	Investigate the education of an ethical, competent professional committed to the society he lives in.	I – assess the level of understanding of scientific, technical, social, economic and financial aspects; II – assess the degree of mastery of logical reasoning in the solution of questions in different contexts; III – verify critical-analytical skills; IV – assess the level of quantitative reasoning; V – verify ability to adopt systematic and holistic perspectives; VI – analyze communication and interaction skills; VII – verify the understanding of technological innovations applied; VIII – assess the perception of ethical conduct.
Competences and Skills Assessed	Skills: read and interpret texts; analyze and criticize information; draw conclusions by induction and/or deduction; establish relations, comparisons and contrasts in different situations; detect contradictions; make valuation choices, assessing consequences; question the reality; and argue coherently. Competences: project intervention actions; propose solutions to problem situations; construct integrative perspectives; elaborate syntheses; administer conflicts and act according to ethical principles.	I – use Accounting terminologies and language; II – practice accounting with a systemic and interdisciplinary view; III – master the identification, recognition, measuring and disclosure process; IV – demonstrate critical-analytic skills, involving verifications, auditing, expertise, arbitrage and quantifications of financial, equity and governmental information; V – demonstrate skills to identify and produce information for the decision process; VI – interpret and apply accounting standards; VII – possess skills to identify users' information need to support the development of information systems; VIII – understand ethical conduct in the practice of accounting activities.
Knowledge Object	Art and culture; technological advances; science, technology and innovation; democracy, ethics and citizenship; ecology and biodiversity; globalization and geopolitics; public policies: education, housing, sanitation, health, transportation, safety, defense, sustainable development; work relations; social accountability: public, private, service sector; sociodiversity: multiculturalism, tolerance, inclusion/ exclusion, gender relations; information and communication technologies; urban and rural life; and violence.	I – Accounting Theory; II – Professional Ethics; III – Financial Accounting; IV – Financial Statement Analysis; V- Cost Accounting and Analysis; VI – Management Accounting and Controllership; VII – Financial Management; VIII – Accounting Applied to the Public Sector; IX – Auditing and Expertise; X – Financial, corporate, occupational and tax legislation; XI – Quantitative methods applied to Accounting; XII – Information systems and technologies.

Source: elaborated by the authors based on Inep (2014).

In this phase, the intent was to identify the reasons for placing the items in each point of the scale, in function of the questions' level of difficulty. In this analysis, the complexity of the objectives, the competences and the knowledge objects was taken into account, including aspects of interdisciplinarity involved in the content of the question. This analysis, together with the creation of the measuring scale and the identification of the "anchor" levels, is presented in the next topic.

4. Analysis of Results

Next, the results of the application of IRT to measure the Accounting students' proficiency (performance) on the Enade test/2012 are presented. It should be highlighted that, out of 35 questions analyzed, eight were excluded due to the very low biserial correlation coefficients, which are inappropriate to estimate the model and can cause distortions in the estimations. The following questions were excluded: 05 and 08, belonging to the general knowledge group; and 10, 19, 21, 28, 29 and 32, belonging to the specific knowledge group. This exclusion is due to the fact that the problems identified in the items appointed did not permit convergence in the estimation of the model using the marginal maximum likelihood method. In other words, the problems the items presented are that big that they make it impossible to calibrate IRT parameters. As the items possess low biserial correlation coefficients, the questions are inefficient to explain the students' performance, as the possibilities of individuals with high or low proficiency levels to answer the item correctly are similar and, therefore, the items do not help to estimate and distinguish the students' performance.

4.1 Estimation of Parameters

Table 2 presents the discrimination (a_i), difficulty (b_i) and casual hit (c_i) parameters of the items assessed, with their respective standard errors (SE). These parameters were estimated in the software BILOG-MG®, using scale 0,1.

As shown in Table 2, the parameters a_i of the items assessed are generally superior to 0.7. Parameter a_i indicates the discrimination of each item – the higher the value of this parameter, the higher its discrimination power. In other words, it indicates the extent to which each item is able to distinguish those individuals who possess the knowledge under evaluation from those who do not. The items with greater discrimination power were items 34, 14 and 11. As shown in Figure 4, these items possess the most inclined item characteristic curves.

Table 2

Estimations of 3PL Item Parameters

Items	*a*	SE	*b*	SE	*c*	SE
1	0.8271	0.0602	0.4967	0.1252	0.2787	0.0322
2	0.5854	0.0314	0.8857	0.1042	0.0929	0.0257
3	0.4705	0.0517	2.0241	0.1600	0.1592	0.0363
4	0.4435	0.0217	0.5318	0.1430	0.0893	0.0271
6	0.4496	0.0397	0.9140	0.2540	0.1740	0.0456
7	0.9238	0.0527	0.6838	0.0732	0.1814	0.0232
11	2.0006	0.0998	1.4364	0.0196	0.2222	0.0051
12	1.6104	0.0707	0.7517	0.0265	0.2123	0.0105
13	1.3361	0.0963	2.3545	0.0571	0.1704	0.0054
14	2.0909	0.2160	2.7792	0.0841	0.1400	0.0025
15	0.9278	0.0791	1.9782	0.0507	0.2586	0.0131
16	0.8736	0.0190	0.3576	0.0241	0.0238	0.0074
17	1.3547	0.1602	3.1178	0.1465	0.1816	0.0043
18	1.4574	0.1153	2.5441	0.0692	0.1679	0.0044
20	0.4851	0.0478	2.2916	0.1151	0.1195	0.0290
22	0.4242	0.0791	4.3488	0.2956	0.1510	0.0275
23	1.3846	0.0663	0.6966	0.0373	0.2553	0.0133
24	0.8050	0.0372	1.3116	0.0429	0.0596	0.0144
25	0.9345	0.0366	-0.2786	0.0876	0.1214	0.0306
26	0.6932	0.0192	0.7049	0.0381	0.0330	0.0101
27	1.2058	0.0566	1.2264	0.0274	0.1281	0.0106
30	1.9193	0.1551	2.5244	0.0613	0.1313	0.0029
31	0.8399	0.1058	3.3597	0.1826	0.1572	0.0094
33	0.5174	0.0838	4.7095	0.3753	0.0834	0.0151
34	2.1004	0.1432	1.9395	0.0321	0.2558	0.0039
35	0.7550	0.0443	1.1123	0.0703	0.1110	0.0219

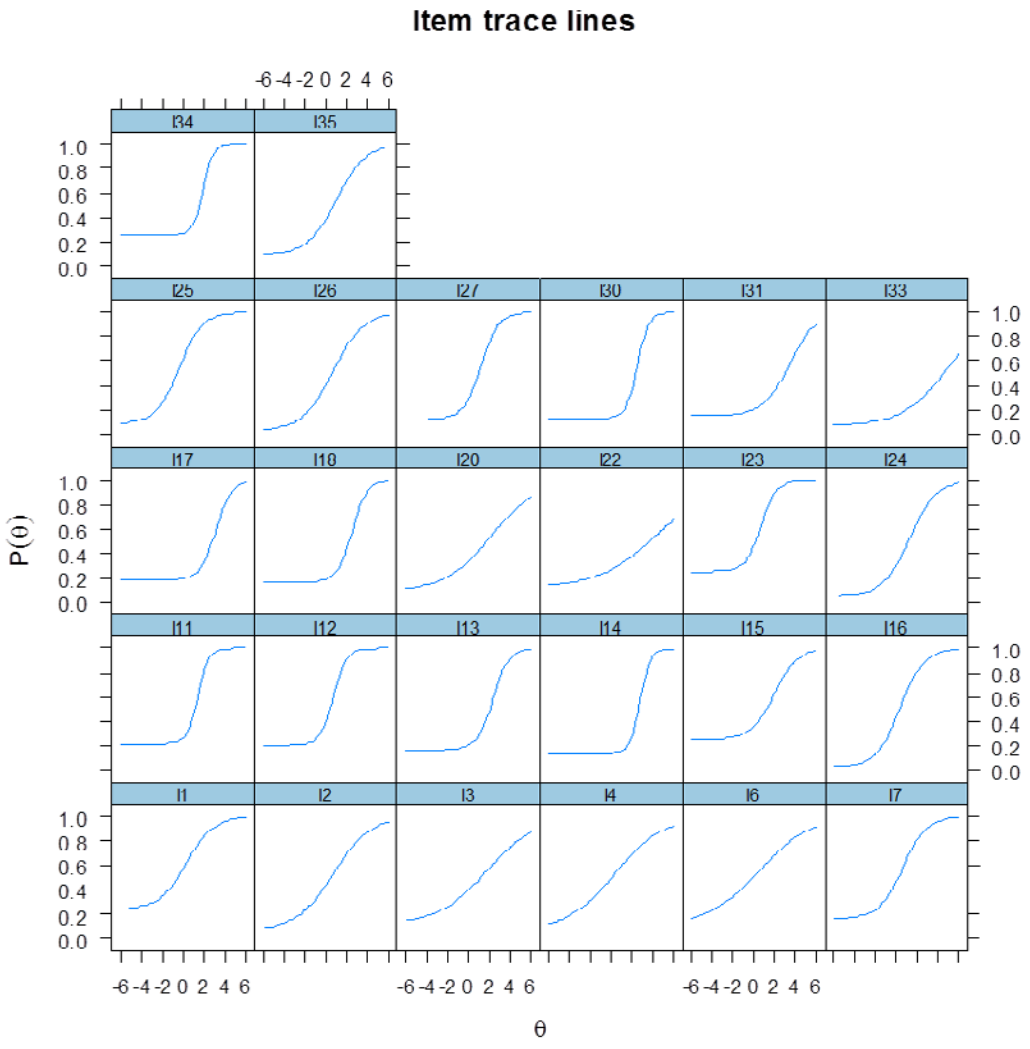
Legend: a = discrimination parameter of IRT; b = difficulty parameter of IRT; c = probability of casual hit according to IRT; SE = Standard Error.
Obs.: The parameters a, b and c presented in the table refer to the grouped item parameters.

Source: elaborated by the authors.

Parameter b_i represents the difficulty of each item – the higher its coefficient, the greater the proficiency the students need to answer the question. According to Table 2, the items with the highest degree of difficulty for the students assessed were 33, 22 and 31. These questions demand students' interdisciplinary knowledge related to Financial Accounting, Accounting Theory, Auditing, Information Systems and Financial and Tax Legislation. The analysis of the questions reveals that the student needs mastery of accounting terminology, critical-analytical skills, a systemic and holistic perspective to interpret and apply the theory and standards inherent in Accounting. On the other hand, the items the students found easier were 25, 16 and 1, which require but basic knowledge on accounting and auditing standards. Item 1 refers to a general knowledge question and demands text reading and interpretation skills. In general, questions 25 and 16, which require specific knowledge, require but the understanding of the text of the professional standard, without demanding critical-analytic skills from the student.

It should be observed that, according to Andrade *et al.* (2000), due to the mathematical definition inherent in the model used, the items are expected to present parameters b_i between -2 and +2 when the scale 0,1 is adopted. Nevertheless, as observed, the items assessed on Enade/2012 present a high degree of difficulty for the students assessed, that is, the test questions greatly exceed the proficiency of the group analyzed.

Figure 4 presents the graphs with the item characteristic curves (ICC). Axis X represents the student's proficiency and axis Y the probability of a correct answer.

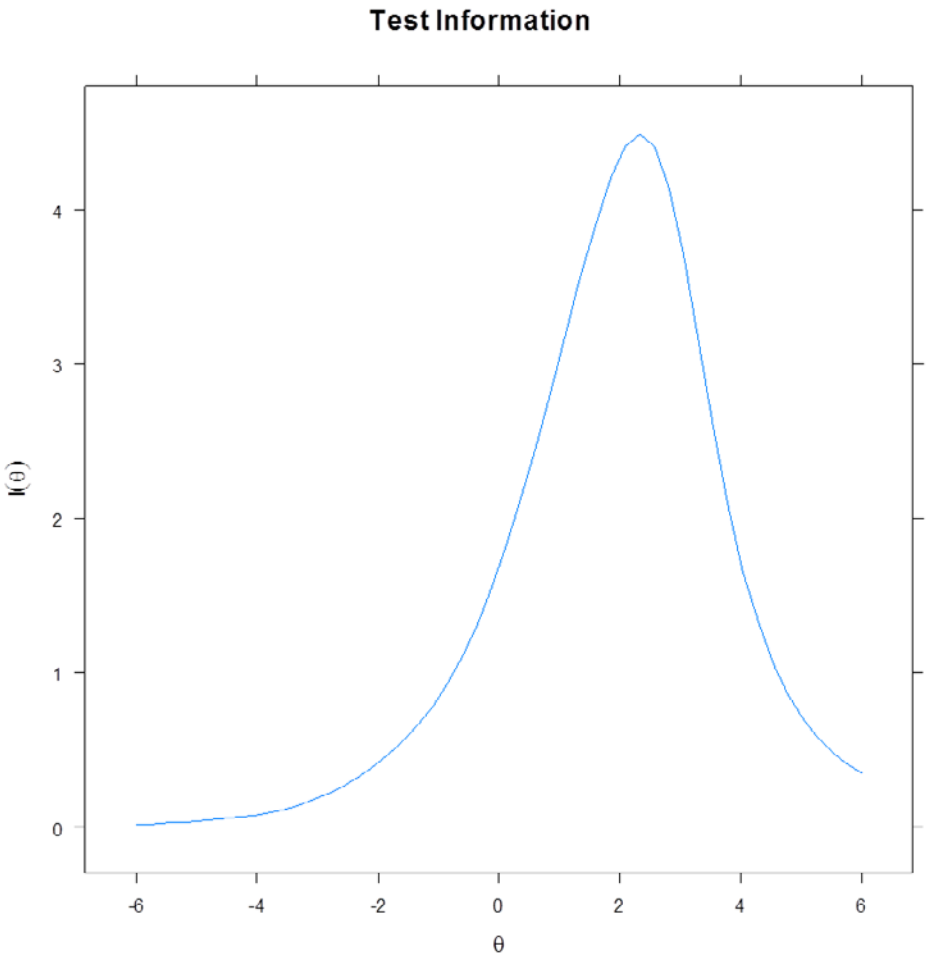


Source: elaborated by the authors.

Figure 4. Item Characteristic Curve (ICC)

Figure 4 displays the variation in the inclination of the item curves. This inclination is related to parameter a_i, that is, the discrimination of the items assessed. As explained in section 3.2.2, these curves express the items' power to discriminate individuals who master the knowledge from those who do not. As it represents the derivative of the inflection point of the curve (ICC), Figure 4 should be considered more in function of the item characteristic curve than in function of the values of θ and P(θ). In other words, items with less steep curves (ICC) or curves closer to straight lines have lesser discrimination power. As a complement to Table 1, items with lower present a less steep ICC, like in the case of items 22, 4, 6, 3, 20 and 33. Oppositely, items with higher a_i parameters, like items 34, 14, 11, 30, 12 and 18, present a steeper ICC.

Figure 5 exhibits the Test Information Function (TIF). This measures consists of the sum of the information functions of all items.

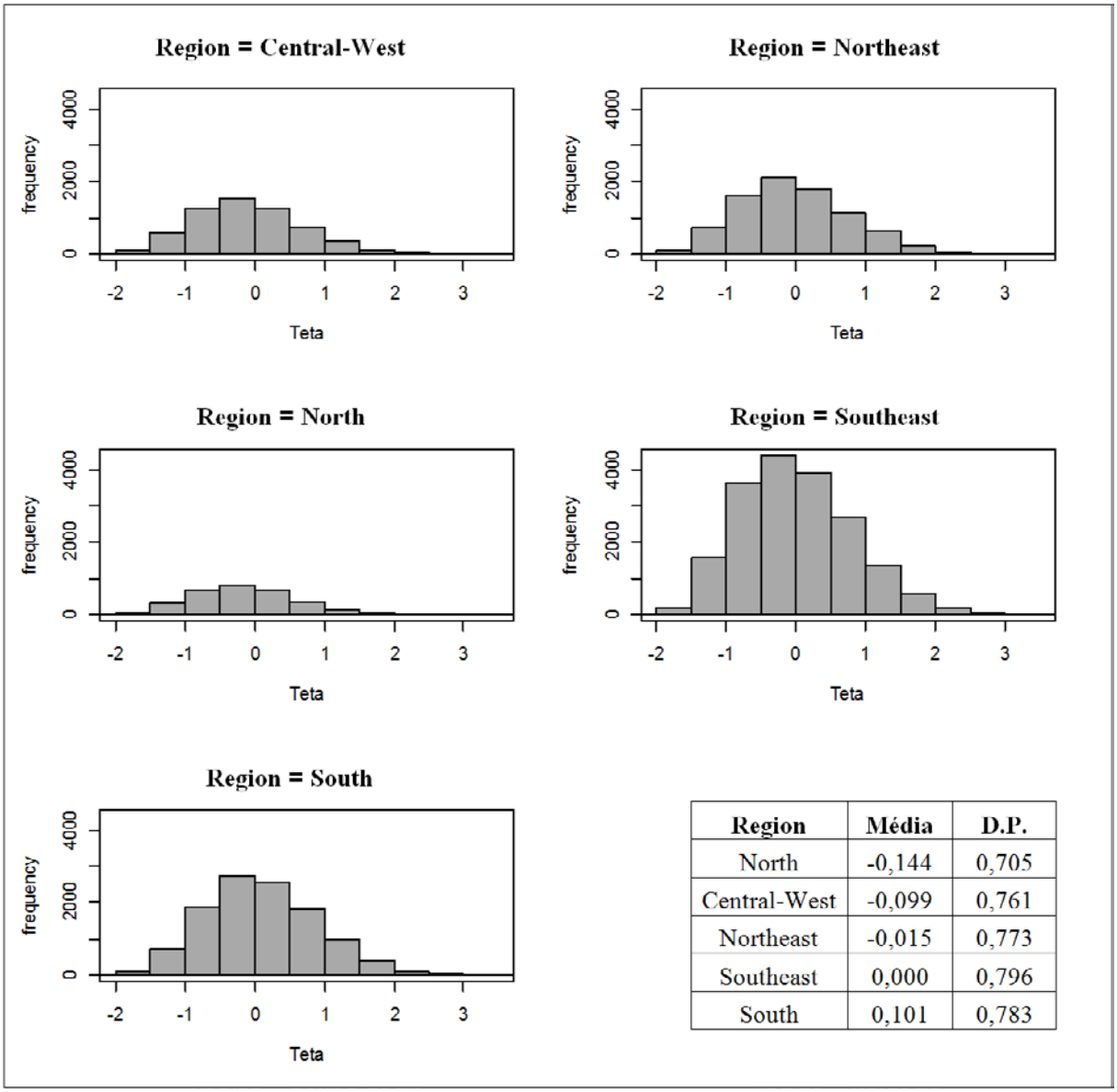


Source: elaborated by the authors.

Figure 5. Test Information Function

As shown in Figure 5, the TIF contain a larger amount of information in the scale interval from 0 to 4. That represents that the Enade/2012 applied to Accounting students represents a high degree of difficulty for the students investigated. It can be observed that only individuals with proficiency scores of two or higher have a high probability of performing well. Thus, it is a more suitable assessment tool to measure the latent trait of students whose proficiency level is close to two.

In Figure 6, the distribution of the students' proficiency level among the regions of the country is displayed. The division per region was adopted in function of the results of Cruz *et al.* (2013), who found significant performance differences among students from different regions on the Enade/2009.



Region	Média	D.P.
North	-0,144	0,705
Central-West	-0,099	0,761
Northeast	-0,015	0,773
Southeast	0,000	0,796
South	0,101	0,783

Source: elaborated by the authors.

Figure 6. Distribution of Proficiency (D.P.) per Region of the Country.

It can be observed in Figure 6 that, independently of the region of Brazil the student belongs to, his/her proficiency tends to score inferior to 0. As observed earlier, the Enade test demonstrated a high degree of difficulty for the group that took the test.

In a complementary analysis of each student group's mean score per region, it is observed that the students from the South show an average proficiency (0.101) slightly higher than in the other regions. The opposite is observed for the North, where the students' average proficiency (-0.144) is slightly lower. These results support the findings by Cruz *et al.* (2013), as the students from the Northeast, North and Central-West present lower performances. In Table 3, the mean proficiency levels of the students are presented, estimated using the 3PL of IRT, in accordance with some characteristics.

Table 3

Proficiency of students per characteristics

Characteristic		n	Mean	SD
Secondary School	Public School	13,371	0.042	0.824
	Private School	33,727	-0.017	0.764
Mother's Education	Higher Education	6,474	0.054	0.855
	Other	40,624	-0.009	0.769
Father's Education	Higher Education	5,666	0.092	0.866
	Other	41,432	-0.013	0.769
Hours of Study Outside Class	More than 3h	15,439	0.096	0.809
	Less than 3h	31,659	-0.047	0.764
Higher Education Institution	Public	7,610	0.184	0.846
		39,488	0.035	0.764

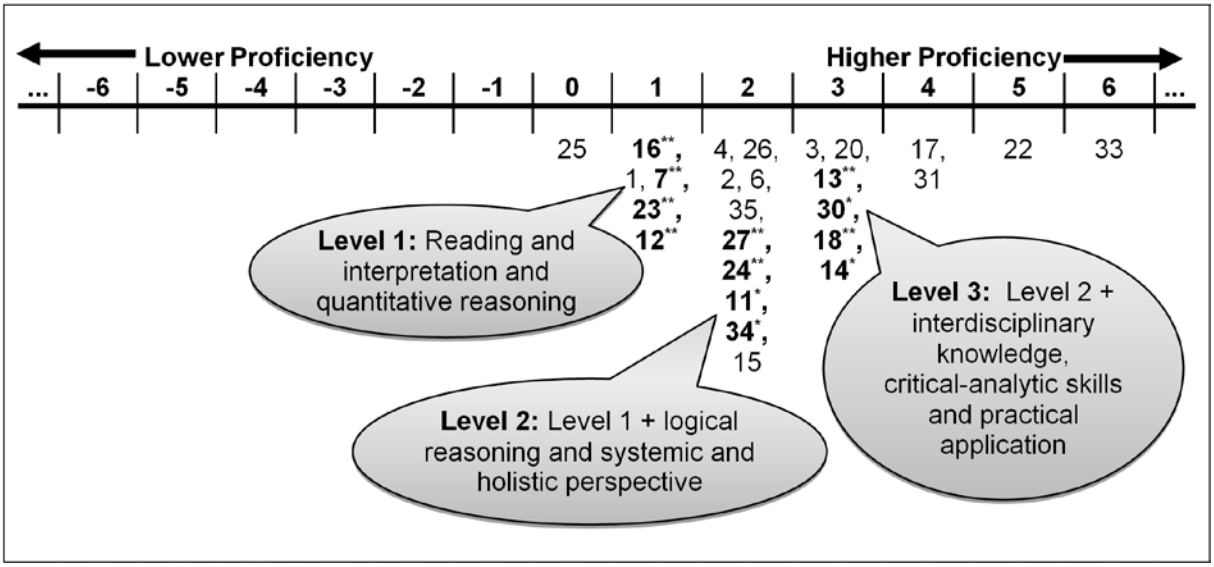
Source: elaborated by the authors.

In Table 3, it is shown that students who took secondary education at a public school, whose mother and father hold a higher education degree, who study more hours outside the classroom and who study at public HEI have slightly higher average proficiencies than students without these characteristics. It should be highlighted that the largest difference between the mean proficiency levels is found between students from public and private HEI. These results support the findings by Souza (2008) and Santos (2012), which were based on previous Enade tests.

In the next topic, the proficiency (performance) scale for the Enade/2012 is presented.

4.2 Proficiency (Performance) scale

The Proficiency scale of the students who took the Enade/2012 is displayed in Figure 7. All items are represented in the scale, but only the items highlighted (in bold) refer to "anchor" or "almost anchor" items.



Legend: * "Anchor" Item; ** "Almost anchor" item.

Source: elaborated by the authors.

Figure 7. Proficiency (performance) scale of Enade/2012 test in Accounting

Levels 1, 2 and 3, presented in Figure 7, were defined according to the "anchor" and "almost anchor" items, identified based on the rules presented in Figure 3. The knowledge and competences summarized at each level were identified according to the knowledge and competences required by the questions of the "anchor" and "almost anchor" levels. The items placed more to the left on the scale, demonstrated in Figure 7, demanded a lower proficiency level from the students. As an example, test question 25 required basic knowledge from the students on the accounting terminologies related to Accounting Theory and Financial Accounting. For the sake of a high probability to correctly answer that question, the students should merely understand the text of the professional standard, without requiring critical-analytic skills.

On the opposite, the items placed more to the right on the scale (Figure 7) required not only the mastery of more complex knowledge, but also critical-analytic skills, with a systemic and holistic perspective. To give an example, question 33 demanded interdisciplinary knowledge between Accounting Theory and Financial Accounting from the students, involving its applications in the identification, recognition, measuring and accounting disclosure process, besides requiring the mastery of accounting terminologies, critical-analytic skills and a systemic and holistic view to interpret and apply the theory and standards inherent in accounting. Question 22 demanded interdisciplinary knowledge between Internal Auditing and Information Systems from the students, involving its applications to the internal controls of the inventory subsystems and requiring the mastery of accounting terminologies, critical-analytic skills and a systemic and holistic view.

To solve question 17, the student should master knowledge on quantitative methods applied to Accounting that contribute to produce information for the decision process and possess critical-analytic skills to analyze the information presented. Question 31 demanded interdisciplinary knowledge involving Financial Accounting and Financial and Tax Legislation. To solve the question, the student should master Financial Accounting knowledge, mainly related to the identification of the amount of profit distributed to the partners or shareholders, in line with financial criteria and in accordance with tax legislation on the assumed profit. In this question, the mastery of accounting terminologies was also required, as well as the interpretation and application of Accounting standards, critical-analytic reasoning and a systemic and holistic perspective.

Based on the estimates measures, three proficiency levels were identified for the Enade/2012. These items are displayed in Table 4. The rules to define these items as "anchor" or "almost anchor" items were presented in topic 3.3 of this article.

Table 4
Classification of items per scale level

Panel A: Classification of Items		
Level	**Items**	**Classification**
Level 1	16, 7, 23 and 12	Almost anchor.
Level 2	11 and 34 27 and 24	Anchor. Almost anchor.
Level 3	30 and 14 13 and 18	Anchor. Almost anchor.

Panel B: Description of Items	
Item	**Description**
	Level 1
16	Requires knowledge about the standards ruling the profession of independent auditor concerning the aspect of professional independence. Involves descriptions of situations that characterize possible threats on the auditor's independence. Despite demanding knowledge of circumstances that affect the independence, requires understanding of the text of the professional standard, without demanding critical-analytic skills.
7	Requires ability to read and interpret texts, with a view to the ability to reach conclusions and/or make inferences based on the use of induction and/or deduction. In general, the question aims to identify whether the student can reach conclusions on a certain theme based on information implicit in the context presented.
23	Involves financial mathematical calculations to calculate the payback for the assessment of project feasibility. As the data are presented in tables, skills like the interpretation and critical analysis of the context the question is presented in are not required.
12	Demands knowledge of cost analysis in decisions involving the choice to purchase or produce goods. Involves a procedure to produce information for (low complexity) decision taking. Requires the mastery of accounting terminologies and the use of quantitative reasoning to identify and verify production costs.
	Nível 2
11	Requires knowledge on cost verification, specifically of the activity-based costing (ABC) model. Is part of the list of procedures to produce information for the decision process. Demands the mastery of accounting terminologies and the use of logical and quantitative reasoning to solve problems with a systemic and holistic view.
34	Requires the calculation of indicators for the analysis of financial statements through the understanding and knowledge of accounting terminology, as well as through the use of quantitative reasoning. Does not require the interpretation of these indicators.
27	Requires the development of a Cost-Volume-Profit (CVP) analysis to identify amounts to be sold by achieving the point of economic equilibrium. Is an auxiliary procedure to produce information for the decision process. The question demands knowledge of accounting terminologies and the use of logical and quantitative reasoning to solve problems with a systemic and holistic view.
24	Demands knowledge on the Statement of Value Added (SVA), specifically regarding its goal and the way the value added is calculated. Requires basic knowledge of accounting terminologies.
	Nível 3
30	Requires interdisciplinary knowledge between the verification of accounting income and taxes through the application of tax legislation to verify the actual profit. Involves the mastery of accounting terminologies, interpretation and application of accounting standards and critical-analytic skills with a systemic and holistic view.
14	Demands knowledge on the accounting identification, recognition, measuring and disclosure process. Involves the mastery of accounting terminologies, interpretation and application of the accounting theory and standards, as well as critical-analytic skills to identify the contextualized phase of the accounting cycle.
13	Demands knowledge of cost-volume-profit analysis to identify the production combination capable of maximizing the company's contribution margin in a context that involves factors that limit the production process. Consists of a procedure to produce information for the decision process and demands the mastery of accounting terminologies and the use of logical and quantitative reasoning to solve problems with a systemic and holistic view.
18	Demands knowledge and skill to verify results through absorption and variable costing models. Requires knowledge of accounting terminologies. Involves critical-analytic analysis with systemic and holistic view through quantitative reasoning for parallel cost verifications.

Source: elaborated by the authors.

As can be seen in Panel B of Table 4, students with Level 1 proficiency are able to perform tasks related to the reading and interpreting of texts and standards, arithmetic calculations and financial mathematics to solve low complexity problems. As we move on to Level 2, the items will demand not only quantitative, but also logical reasoning from the students. In addition to reading and interpretation of texts and standards, at this level, systemic and holistic perspectives related to the production of information for decision making are also necessary.

At Level 2, item 24 on the Statement of Value Added (SVA) should be highlighted. Although it does not require highly complex skills and knowledge, positioning the item at that level of the scale suggests the need for greater attention to the course curricula. Due to the recent convergence process and curriculum issues, the study of this statement may have been compromised. As can be seen, this item has been ranked as of medium complexity, but only requires general knowledge on the purpose and the SVA components, without further requirements, like in the cases of other questions at this level which required logical reasoning and a systemic and holistic perspective.

At Level 3 (Table 4), besides the mastery of knowledge and skills required in previous levels, such as quantitative and logical reasoning and systemic and holistic perspectives, items at this level require interdisciplinary knowledge of accounting contents and critical and analytical skills. At Level 3, reading and Interpretation of texts and standards are the starting point for practical application of knowledge mastered by the student. As expected, the domain of knowledge, skills and competencies are cumulative from the displayed levels. The range includes initially simpler domains, requiring, as the level increases, more complex areas of students.

In Table 5, the amounts of students classified according to the proficiency levels set out in Figure 7 are presented.

Table 5
Classification of students per proficiency range

Ranges	Number of Students	Frequency
Inferior to 1	35,120	75%
Level 1	10,169	22%
Level 2	1,724	4%
Level 3	85	0%
Total	**47,098**	**100%**

Source: elaborated by the authors.

Based on the study of the students' individual proficiency, the results show that, out of 47,098 students, 35,120 (75%) ranked below level 1, or had a proficiency level represented by IRT θ lower than the first range (Level 1) of the proficiency scale (Figure 7), which means that most students are not even able to master the knowledge and minimum skills required for the examination; 10,169 (22%) students are proficient at Level 1; only 1,724 students (1%) at Level 2, and only 85 students master the knowledge required at Level 3.

Similar to studies by Scher et al. (2014), Coelho et al. (2014) and Primi et al. (2011), the application of IRT to the Enade test in Accounting was feasible. Specifically, as performed by Primi et al. (2011), it was possible to establish performance ranges to segregate the types of knowledge were segregated, skills and competencies required to fit students into each range.

5. Conclusion and Suggestions for Future Research

This study aimed to measure the performance (proficiency) of Accounting students on Enade, using IRT. Based on the estimation by means of 3PL, deriving from IRT, the proficiency of Accounting students who took the Enade / 2012 was measured and a standardized scale was created.

The analysis of the items (questions) showed evidence of the test's ability to measure the proficiency of students at different levels of cognitive mastery required by the test. In this context, the IRT demonstrated ability to capture the distribution of Accounting students' proficiency across the levels required by the test.

The survey results showed that the Enade items represent a high degree of difficulty for the group that performed the test. Both in the regional analysis, or based on other characteristics (e.g. parental education, time to study or undergraduate course at public or private HEI), the students' proficiency tends to be low, i.e. below 0, considering the 0.1 scale.

Most students who performed the test did not even show the cognitive mastery of scale Level 1. This result corroborates the low performance of students appointed by the INEP report (2014) for this test and specifically appoints in what aspects and knowledge learning weaknesses can be found. Unfortunately, this negative performance was also observed in previous accounting tests (Santos, 2012). As these assessments are intended to monitor the situation of undergraduate education, it is important that action is taken, involving public policies to revert this situation. In this case, it should be investigated whether the cause of the degree of difficulty of the test items is related to the complexity of the questions or to the inefficient application of educational resources and public policies adopted in HEIs, making the teaching-learning process unsatisfactory.

It is also worth noting that, while studies based on CTT address the issue of performance in an aggregated manner, the analysis by means of IRT permitted the timely identification of the knowledge, abilities and skills required at each level of the scale. When dismembering the proficiency levels, demonstrating the required knowledge, punctual actions of teachers, HEI and the respective authorities are possible, acting on specific aspects in which learning disabilities were demonstrated.

Given the benefits mentioned and presented by IRT in this study, further research related to the determinants of Accounting students' performance could adopt measures provided by models based on IRT as a base for student performance measuring. This analysis would make it possible to study determinants based on different levels of proficiency, contributing to the advancement of studies related to this area.

6. References

Andrade, D. F., Tavares, H. R., & Valle, R. C. (2000). *Teoria da Resposta ao Item:* conceitos e aplicações. São Paulo: Associação Brasileira de Estatística (ABE).

Bartlett, S., Peel, M. J., & Pendlebury, M. (1993). From fresher to finalist: a three year analysis of student performance on an accounting degree programme. *Accounting Education*, *2*(2), pp. 111. http://dx.doi.org/10.1080/09639289300000013

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *The journal of political economy, 70*(5), pp. 9-49. http://dx.doi.org/10.1086/258724

Birnbaum, A. (1968). Some Latent Trait Models and Their Use in Infering an Examinee's Ability. In Lord, F.M., & Novick, M.R. (Ed.) *Statistical Theories of Mental Test Scores. Reading*, MA: Addison-Wesley.

Blaug, M. (1992). *The methodology of economics: Or, how economists explain.* Cambridge University Press.

Bock, R. D. (1997). A brief history of Item Response Theory. *Educational Measurement:* Issues and Practice, *16*(4), pp. 21-33. http://dx.doi.org/10.1111/j.1745-3992.1997.tb00605.x

Boomsma, A., Van Duijn, M. A. J., & Snijders, T. A. B. (2000). Essays on item response theory. *Lecture Notes in Statistics* (Springer-Verlag), pp. 157. New York: Springer.

Byrne, M., & Flood, B. (2008). Examining the relationships among background variables and academic performance of first year accounting students at an Irish University. *Journal of Accounting Education, 26*(4), pp. 202-212. http://dx.doi.org/10.1016/j.jaccedu.2009.02.001

Clinton, B. D., & Kohlmeyer, J. M. (2005). The effects of group quizzes on performance and motivation to learn: Two experiments in cooperative learning. *Journal of Accounting Education, 23*(2), pp. 96-116. http://dx.doi.org/10.1016/j.jaccedu.2005.06.001

Coelho, E. C., Ribeiro Junior, P. J. & Bonat, W. H. (2014). Exame nacional de desenvolvimento de estudantes de estatística-desafios e perspectivas pela TRI. *Revista da Estatística da Universidade Federal de Ouro Preto, 3*(2), pp. 323-337.

Cruz, A. J., Nossa, V., Balassiano, M., & Teixeira, A. (2013). Desempenho dos Alunos no Enade de 2009: um estudo empírico a partir do conteúdo curricular dos cursos de Ciências Contábeis no Brasil. *Advances in Scientific and Applied Accounting, 6*(2), pp. 178-203. http://dx.doi.org/10.14392/asaa.2013060203

Cunha, J.V.A., Cornachione Junior, E.B. & Martins, G.A. (2010, Mai./Jun.). Doutores em Ciências Contábeis: Análise sob a Óptica da Teoria do Capital Humano. *RAC,* Curitiba, *14*(3), pp. 532-557. http://dx.doi.org/10.1590/s1415-65552010000300009

Davidson, R. A. (2003). Relationship of study approach and exam performance. *Journal of Accounting Education, 20*(1), pp. 29-44. http://dx.doi.org/10.1016/s0748-5751(01)00025-2

Edelen, M. O. & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research, 16*(1), pp. 5-18. http://dx.doi.org/10.1007/s11136-007-9198-0

Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of human Resources, 14*(3), pp. 351-388. http://dx.doi.org/10.2307/145575

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP. (2014). *Relatório de Síntese, Ciências Contábeis - ENADE 2012*. Recuperado de http://portal.inep.gov.br/web/guest/enade/relatorio-sintese-2012.

Lopes, F. L. & Vendramini, C. M. M. (2015). Propriedades psicométricas das provas de pedagogia do ENADE via TRI. *Avaliação: Revista da Avaliação da Educação Superior, 20*(1), pp. 27-47.

Lord, F. M. (1952). A theory of test scores (No. 7). *Psychometric Monograph*.

Miranda, G. J. (2011). *Relações entre as qualificações do professor e o desempenho discente nos cursos de graduação em contabilidade no Brasil.* Tese de Doutorado em Controladoria e Contabilidade, Faculdade de Economia, Administração e Contabilidade (USP), São Paulo, SP, Brasil. Recuperado de http://www.teses.usp.br/teses/disponiveis/12/12136/tde-16032012-190355/pt-br.php.

Moreira, F. J., Jr (2010). Aplicações da teoria da resposta ao item (TRI) no Brasil. *Rev. Bras. Biom, 28*(4), pp. 137-170.

Primi, R., Hutz, C. S. & Silva, M. C. R. (2011). A prova do ENADE de psicologia 2006: concepção, construção e análise psicométrica da prova. *Avaliação Psicológica, 10*(3), pp. 271-294.

R Development Core Team. (2012). *R:* A Language and Environment for Statistical Computing. R Foun-dation for Statistical Computing, Vienna, Austria.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests.* Copenhagen: Danish Institute for Educational Research.

Rasera, J. B. (1999). L'économie de l'éducation et la question du développement. In: Paul, J.-J. (Ed.) *Administrer, gérer,* évaluer *les systèmes* éducatifs. Paris: ESF.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika monograph supplement*.

Santos, N. A. (2012). *Determinantes do desempenho acadêmico dos alunos dos cursos de ciências contábeis.* Tese de Doutorado em Controladoria e Contabilidade, Faculdade de Economia, Administração e Contabilidade (USP), São Paulo, SP, Brasil. Recuperado de http://www.teses.usp.br/teses/disponiveis/12/12136/tde-11062012-164530/en.php.

Santos, N. D. A., Cunha, J. V. A. D., & Cornachione Junior., E. B. (2009). Análise do Desempenho dos Cursos de Ciências Contábeis do Estado de Minas Gerais no ENADE/2006. *Anais do Congresso da Associação Nacional dos Programas de Pós Graduação em Ciências Contábeis – ANPCONT,* IAAER-ANPCONT, São Paulo, SP, Brasil, 3.

Scher, V. T., Moreira, F. J., Jr., Correa, A. C., Schuch Junior, V. F., Andrade, D. F., & Bortolotti, S. L. V. (2014). Uma aplicação da teoria da resposta ao item na avaliação do ENADE do Curso de Administração. *Anais do XIV Colóquio Internacional de Gestão Universitária – CIGU*, Florianópolis, SC, Brasil, 14.

Schultz, T. W. (1960). Capital formation by education. *The journal of political economy, 68*(6), pp. 571-583.

Souza, E. S. (2008). *ENADE 2006: Determinantes do Desempenho dos Cursos de Ciências Contábeis.* Dissertação de Mestrado em Ciências Contábeis, Programa Multiinstitucional e Inter-Regional de Pós-Graduação em Ciências Contábeis, Universidade de Brasília/UFPB/UFRN, Brasília, DF, Brasil. Recuperado de http://repositorio.unb.br/handle/ 10482/3949.

Van Der Linden, W. J. & Hambleton, R. K. (1997). *Handbook of Modern Item Response Theory.* New York: Spring-Verlag.

Waltenberg, F. D. (2006, jan./abr.) Teorias econômicas de oferta de educação: evolução histórica, estado atual e perspectivas. *Educação e Pesquisa, 32*(1), pp. 117-136. http://dx.doi.org/10.1590/s1517-97022006000100008

Wright, B. D. (1968, February, 11-13). Sample-free Test Calibration and Person Measurement. *Paper Presented At The National Seminar On Adult Education Research (Chicago).*