

## Analysis of distractors: an interpretation by the nominal response model of Enade 2017 items applied to the Degree in Mathematics

Análise de distratores: uma interpretação pelo modelo de resposta nominal dos itens do Enade 2017 aplicado à Licenciatura em Matemática

Análisis de distratores: una interpretación por el modelo de respuesta nominal de los ítems de Enade 2017 aplicados a la Licenciatura en Matemáticas

**Sandra Cristina Martini Rostirola** - Universidade do Estado de Santa Catarina | Instituto Federal Catarinense | Videira | Santa Catarina | Brasil. E-mail: [sandra.rostirola@ifc.edu.br](mailto:sandra.rostirola@ifc.edu.br) | Orcid: <https://orcid.org/0000-0003-2482-6117>

**Elisa Henning**- Universidade do Estado de Santa Catarina | Joinville | Santa Catarina | Brasil. E-mail: [elisa.henning@udesc.br](mailto:elisa.henning@udesc.br) | Orcid: <https://orcid.org/0000-0002-7754-9451>

**Ivanete Zuchi Siple** - Universidade do Estado de Santa Catarina | Joinville | Santa Catarina | Brasil. E-mail: [ivanete.siple@udesc.br](mailto:ivanete.siple@udesc.br) | Orcid: <https://orcid.org/0000-0002-8640-1336>

**Abstract:** Multiple-choice item-based tests often use Item Response Theory for parameter estimation within dichotomous models. However, the Nominal Response Model allows the study of polytomous items, considering all the information obtained from the analysis of alternatives chosen by respondents. In this regard, the present research aims to interpret the content and alternatives of the Enade 2017 items related to Mathematics Teaching courses to contribute to a deeper edumetric analysis of the said test. It is a quanti-qualitative approach that concurrently employs the Nominal Response Model and the perspective of mathematics error analysis. The results indicate that the Enade 2017 test for the Mathematics Teaching course contained items with weaknesses in technical aspects in terms of clarity, relevance, or difficulty, requiring review by experts in the field. Furthermore, the analysis of these items helped clarify some of the respondents' reasoning.

**Keywords:** large scale assessment; nominal response model; degree in mathematics.

DOI: <http://dx.doi.org/10.1590/S1414-40772023000100043>

**Resumo:** Testes baseados em itens de múltipla escolha frequentemente utilizam a Teoria da Resposta ao Item para a estimação de parâmetros dentro de modelos dicotômicos. Contudo, o Modelo de Resposta Nominal permite estudar itens politômicos, considerando toda a informação obtida na análise das alternativas escolhidas pelos respondentes. Nesse sentido, a presente pesquisa objetiva interpretar o conteúdo e as alternativas dos itens do Enade 2017 referentes aos cursos de Licenciatura em Matemática, a fim de contribuir para o aprofundamento da análise edumétrica do referido teste. Trata-se de uma abordagem quantiquantitativa que utiliza, concomitantemente, o Modelo de Resposta Nominal e a perspectiva da análise de erros em matemática. Os resultados indicam que o teste Enade 2017 para o curso de Licenciatura em Matemática continha itens com fragilidades em aspectos técnicos em termos de clareza, pertinência ou dificuldade, com necessidade de revisão por experts da área. Além disso, a análise desses itens permitiu explicar alguns raciocínios dos respondentes.

**Palavras-chave:** avaliação em larga escala; modelo de resposta nominal; licenciatura em matemática.

**Resumen:** Las pruebas basadas en ítems de opción múltiple a menudo utilizan la Teoría de la Respuesta al Ítem para la estimación de parámetros dentro de modelos dicotómicos. Sin embargo, el Modelo de Respuesta Nominal permite estudiar ítems politómicos, considerando toda la información obtenida en el análisis de las alternativas elegidas por los encuestados. En este sentido, la presente investigación tiene como objetivo interpretar el contenido y las alternativas de los ítems del Enade 2017 relacionados con los cursos de Licenciatura en Matemáticas, con el fin de contribuir a un análisis edumétrico más profundo de dicho examen. Se trata de un enfoque cuanti-cualitativo que utiliza simultáneamente el Modelo de Respuesta Nominal y la perspectiva del análisis de errores en matemáticas. Los resultados indican que el examen Enade 2017 para el curso de Licenciatura en Matemáticas contenía ítems con debilidades en aspectos técnicos en términos de claridad, relevancia o dificultad, que requerían revisión por parte de expertos en el campo. Además, el análisis de estos ítems permitió explicar algunos razonamientos de los encuestados.

**Palavras clave:** evaluación a gran escala; modelo de respuesta nominal; licenciatura en matemáticas.

## 1 Introduction

The Item Response Theory (IRT) provides a range of tools to measure a given latent trait, which have proven accurate in the applications of the Basic Education Assessment System (SAEB, from the Portuguese term “Sistema de Avaliação da Educação Básica”) and the National Secondary Education Examination (ENEM, from “Exame Nacional do Ensino Médio”). However, the National Student Performance Examination (Enade, from the Brazilian Portuguese “Exame Nacional do Desempenho de Estudantes”) still does not employ its methods, adopting the Classical Test Theory (CTT). The IRT approach uses Bock’s Nominal Response Model (NRM) (1972), which considers all response categories of a question, using all the information contained in each participant’s responses.

Pinheiro, Costa and Cruz (2010) explain that, in the NRM, the most frequently selected alternative with the highest latent trait is potentially the correct one, while the others consist of distractors. If that is not the case, an answer sheet error, typos on the answers, or poor question formulation can be considered, since the evaluated latent trait must advance according to the answers. Considering this reasoning with the use of the NRM, a certain test can be edumetrically<sup>1</sup> analyzed, highlighting its strengths and weaknesses, and understanding the reasoning of the respondents.

Thus, the use of this model can aid the interpretation of the test as a whole, since each of the alternatives is considered in the analysis of a question. This may represent a qualitative gain for a test such as Enade, which evaluates competences of Brazilian higher education students who are about to graduate.

---

<sup>1</sup> Edumetry (or its variants: edumetric, edumetrically) is the term used to express the in-depth analysis of the quality of a given assessment, considering its validity and reliability.

This work intends to interpret the contents and alternatives of Enade 2017's questions regarding the undergraduate program in Mathematics, in order to contribute to the deepening of its edumetric analysis. In the adopted quantitative-qualitative methodology, we present a study that, beyond estimating the questions' quantitative parameters, observes their structure from a qualitative perspective, with the detailing of possible reasonings from the participants. In this sense, we argue for the need to understand that the errors can carry information that is often overlooked when considering only the right and wrong dichotomy to form a score. To this end, after the introduction, the theoretical foundations are presented, followed by the detailing of the research method, the data and obtained results' analysis and the final considerations.

## 2 Literature review

Discussions about educational evaluation are extremely important, once it allows to assess, based on concrete data, the quality of education provided to the population. According to Lopes and Vendramini (2015), the primacy of these discussions lies on the need for substantially effective teaching, given its effects on the most diverse sectors and layers that make up the socio-political and economic organization of a country.

IRT has become one of the main tools for test analysis, especially to those from the sphere of educational evaluation for the last undergraduate years, as it allows for an accurate measurement of a latent trait in uni and multidimensional terms. However, the most common research studies and applications suggest the use of hit-or-miss patterns, enabling a dichotomous interpretation of the test's questions (Pinheiro; Costa; Cruz, 2010; Reise *et al.*, 2023; Smith; Bendjilali, 2022). This interpretation assumes the application of a model for dichotomous items. As with Logistic Regression models (Hosmer; Lemeshow, 2000), this binary model can be adapted to cases with more than

two categories, except for the differences between regression analysis and the IRT. Thus, polytomous items can be analyzed, with an ordinal or nominal scale. The measurement scale is nominal if the categories are purely qualitative and there is no natural order (Agresti; Kateri, 2014; Kutner; Nachtsheim; Netter; Li, 2004). In this context, Bock's Nominal Response Model (1972) represents a possibility to measure multiple-choice items considered in a polytomous way.

Bock (1972) states that, through the NRM, the probability of an individual selecting an answer alternative  $k$  of a certain question  $i$ , as a function of the difficulty parameters ( $c$ , in the NRM notation) and discrimination ( $a$ ) associated with the subject's latent trait, can be measured through Equation (1):

$$P(u_i = x|\theta; \mathbf{a}; \mathbf{c}) = \frac{e^{a_x(\theta)+c_x}}{\sum_{k=1,m} e^{a_{ik}(\theta)+c_{ik}}} \quad \text{- Equation (1)}$$

Technically, the parameter estimation procedure through NRM does not require the insertion of the correct answers into the application procedures, as with the logistic models of one, two or three parameters. Hence, the distractors, which are conceptualized as different alternatives to the correct answer (Haladyna; Downing; Rodriguez, 2002), can provide some information, since the distance between the respondent's answer and the correct alternative leads to assumptions regarding the skill to be measured (Pinheiro; Costa; Cruz, 2010). Thus, the respondent's error does not simply yield a reduced score, but provides data with information about the question elaboration and can lead to the understanding of the respondent's development level. Stewart *et al.* (2021) show that a skill trait is consistently associated with correct thinking, while other skill dimensions represent specific modes of incorrect thinking.

An important study converging to this research work was made by Thissen, Steinberg and Fitzpatrick (1989), who, initially, explain the evolution of Bock's model (1972) regarding the overcoming of some technical difficulties while estimating parameters. Later, they carry out the analysis of multiple-choice questions from a large scale assessment considering the NRM and detailing the distractor alternatives, evidencing possible reasoning, skill levels and chance hits related to each question.

Furthermore, studies such as those by Smith and Bendjilali (2022) and Zhang *et al.* (2021) explain that the NRM latent trait measurement incorporates which incorrect answers the respondents selected, thus acknowledging that different incorrect answers can indicate different levels of understanding. These studies present an alternative to estimate parameters and compute scores in large-scale assessments, which can be a useful tool to analyze the results themselves regarding elaboration errors, or even the so-called 'tricky questions', which have the conceptual function of distracting test respondents. Thus, in the analysis, the conclusions by Beltrão and Mandarino (2023) are considered, which state that the question format influences the solving difficulty and that some of them require a more elaborate cognitive level than others.

For the respondents' errors analysis and the interpretation of the possible reasoning used in the resolutions, the studies by Allevato (2004), Cazorla (2002), Cury (2013), Cury and Cassol (2004) and Viali and Cury (2009) were used as our basis. The vision of Krutetskii (1976), who criticizes psychometrics when used just to gauge measures for right and wrong, without considering the underlying cognitive process, was also incorporated into our work. In this sense, we argue for the need to understand that the errors can carry information that is often overlooked when considering only the right and wrong dichotomy to form a score. The next section explains the method used to analyze Enade 2017's questions.

### 3 Methodology

The quantitative-qualitative approach analyzes Enade 2017's test for the undergraduate program in Mathematics using IRT's Nominal Response Model, which considers the answers chosen by a respondent in a multiple-choice test to estimate the parameters.

Like so, this research sets out to perform Enade 2017's edometrics, which allows for an effective analysis of the items and deepens the understanding of the respondents' reasoning development (Pereira; Oliveira; Tinoca, 2010). In this sense, both the correctly-answered questions and the missed ones bring relevant information.

The data were collected from Enade 2017's microdata provided by Brazil's Ministry of Education and Culture (MEC<sup>2</sup>, from Brazilian Portuguese "Ministério da Educação e Cultura"). The sample considered 10 869 participants from undergraduate programs in Mathematics who answered at least one question from the specific knowledge test.

The Enade 2017 test consists of 40 questions, 5 of which are open-ended and 35 are multiple-choice. The multiple-choice part had 8 general-knowledge and 27 specific questions, which were organized with five alternatives (A, B, C, D, E), of which only one was correct. Each question of the multiple-choice test was named, for this analysis, with the capital letter "I" and its respective number from the test. The participants' answer pattern (alternatives chosen) for the 35 multiple-choice questions was organized in a spreadsheet and analyzed through NRM using the R - Rstudio software interface (R Core Team, 2022) and the packages Mirt (Chalmers, 2016) and Psych (Revelle, 2023). The parameters for each response category were estimated, and the data was complemented using the CTT.

---

<sup>2</sup> From the website: <http://www.inep.gov.br>

Each alternative's choosing proportion was also calculated for each item, thus considering the correct answer and the distractors (other alternatives present in the question). According to studies by Martins and Beck (2018), distractors can be classified according to the choosing proportion. A strong distractor has a choosing percentage from 31 % to 80 %, a moderate from 21 % to 30 % and a weak of 20 % or less.

Like so, the discrepant questions were characterized by observing the graph curves and their quantitative parameters, and their edumetric analysis was carried out, which includes the interpretation of the respondent's reasoning in the light of the specific literature in the area of error analysis in mathematics. This allowed for an in-depth view of the questions' contents and the proposal of improvements to the test, considering its importance in assessing the skills of the senior and recently graduated students from undergraduate programs in Mathematics in Brazil, as explained in the next section.

## 4 Results and discussions

### 4.1 Proportion of respondents and parameter estimation

Table 01 shows data on the proportion of respondents for each alternative and for the unanswered questions (NA). The proportions regarding the correct answer are shown in blue and the occurrence of distractors that have a higher choosing proportion than the correct answer is highlighted in red.

Table 01 - Proportion of answer distribution per alternative and no-answer (NA)

(in blue, the correct answer, in red, the distractor with a higher probability than the correct answer)

I (*)	Alternatives						I (*)	Alternatives					
	NA (**)	A	B	C	D	E		NA (**)	A	B	C	D	E
1	0.4	9.2	30.2	22.9	28.1	9.1	19	0.7	18.9	23.7	15.0	27.4	14.2
2	0.3	11.2	6.8	53.3	14.3	14	20	0.6	10.3	34.0	23.3	20.0	11.7
3	1.1	19.5	35.8	18.3	16.7	8.6	21	0.7	12.7	19.4	22.2	29.2	15.7
4	0.4	6.4	56.7	4.0	13.4	19.0	22	0.9	12.5	23.8	17.5	15.0	30.2
5	0.4	2.3	11.6	53.5	9.7	22.5	23	0.8	39.7	13.8	14.4	9.1	22.3
6	0.5	7.1	5.1	6.5	4.4	76.5	24	1.0	20.1	22.9	18.1	12.7	25.3
7	0.5	33	27.0	12.2	15.2	12.1	25	0.8	32.3	23.7	19.3	14.2	9.6
8	0.6	4.5	12.4	13.4	43.3	25.9	26	0.7	11.6	7.6	60.00	9.5	10.7
9	0.5	20.4	18.5	25.3	10.5	24.9	27	0.6	16.3	12.0	11.6	14.1	45.4
10	0.5	54.9	15.4	11.5	11.4	6.3	28	0.7	6.4	7.6	41.2	13.2	31.0
11	0.7	16.4	13.9	30.7	21.6	16.7	29	0.5	6.2	26.7	34.6	11.7	20.4
12	0.5	14.8	32.4	23.0	19.9	9.5	30	0.7	14.0	43.9	9.6	16.0	15.8
13	0.6	13.2	16.7	20.8	38.4	10.4	31	0.8	5.5	15.2	15.9	25.4	37.3
14	0.7	12.8	17.6	26.4	26.8	15.7	32	0.7	41.1	21.6	7.3	15.2	14.1
15	0.5	24.9	20.7	12.9	25.7	15.4	33	1.1	43.4	9.5	8.9	29.6	7.4
16	0.4	27.4	13.9	32.3	11.1	14.8	34	0.9	14.4	19.7	17.7	34.7	12.6
17	0.7	25.2	18.7	27.0	18	10.5	35	0.8	9.5	7.2	9.2	65.3	7.9
18	0.9	17.5	21.9	22.5	20.9	16.2	-	-	-	-	-	-	-

\*: Question name; \*\*NA: No answer.

Source: prepared by the author (2023).

In general, the questions present the highest percentage of respondents in the alternative representing the correct answer, as Table 01 shows. However, items: I1, I14, I16, I18, I20, I23 and I24 have a higher percentage of respondents in a distractor. The distractors shown as the highest choosing proportion in Table 01 are classified as moderate (I14, I18 and I24) to strong (I1, I16, I20 and I23).

The parameters regarding each of the possible answer categories (A, B, C, D, E) were estimated through the NRM, yielding probabilities described as  $a_1, a_2, a_3, a_4, a_5$  on Table 02. Positive numbers indicate more likely responses and negative numbers indicate values associated with less likely responses (Thissen; Steinberg; Fitzpatrick, 1989).

Table 02 - Parameter estimation per alternative from questions obtained through NRM

<b>I</b>	<b><math>a_1</math></b>	<b><math>a_2</math></b>	<b><math>a_3</math></b>	<b><math>a_4</math></b>	<b><math>a_5</math></b>	<b>I</b>	<b><math>a_1</math></b>	<b><math>a_2</math></b>	<b><math>a_3</math></b>	<b><math>a_4</math></b>	<b><math>a_5</math></b>
<b>I1</b>	0.208	-0.324	0.523	-0.223	-0.183	<b>I19</b>	-0.133	0.171	-0.2	0.129	0.033
<b>I2</b>	-0.09	-0.259	0.201	0.069	0.079	<b>I20</b>	-0.088	-0.044	0.27	-0.2	0.062
<b>I3</b>	-0.012	0.300	-0.269	-0.163	0.144	<b>I21</b>	-0.055	0.094	-0.159	-0.005	0.126
<b>I4</b>	-0.28	0.844	-0.546	-0.047	0.028	<b>I22</b>	-0.098	-0.138	-0.222	-0.147	0.6
<b>I5</b>	-0.454	0.01	0.594	-0.087	-0.063	<b>I23</b>	0.202	-0.182	-0.067	-0.35	0.396
<b>I6</b>	-0.178	-0.229	-0.422	-0.105	0.93	<b>I24</b>	-0.042	0.316	-0.177	-0.207	0.111
<b>I7</b>	0.305	-0.04	-0.289	0.273	-0.25	<b>I25</b>	0.41	0.235	0.199	-0.244	-0.203
<b>I8</b>	-0.458	-0.131	-0.049	0.51	0.127	<b>I26</b>	-0.18	-0.418	0.793	-0.269	0.073
<b>I9</b>	-0.013	-0.2	0.261	-0.407	0.359	<b>I27</b>	0.008	-0.316	-0.282	-0.068	0.658
<b>I10</b>	0.654	-0.015	-0.327	-0.263	0.04	<b>I28</b>	-0.466	-0.632	0.66	-0.078	0.516
<b>I11</b>	-0.305	-0.36	0.422	0.008	0.235	<b>I29</b>	-0.662	0.075	0.46	-0.118	0.244
<b>I12</b>	-0.141	0.572	-0.269	-0.257	0.095	<b>I30</b>	-0.212	0.72	-0.596	0.001	0.087
<b>I13</b>	-0.524	-0.583	0.23	0.595	0.282	<b>I31</b>	-0.466	-0.306	0.1	0.146	0.52
<b>I14</b>	0.038	0.158	-0.113	0.014	-0.097	<b>I32</b>	0.607	0.101	-0.527	-0.333	0.152
<b>I15</b>	-0.08	-0.331	-0.115	0.393	0.133	<b>I33</b>	0.286	-0.563	-0.434	0.207	0.504
<b>I16</b>	0.125	-0.023	-0.11	0.052	-0.045	<b>I34</b>	-0.139	-0.066	-0.21	0.358	0.057
<b>I17</b>	-0.014	0.107	0.287	-0.068	-0.313	<b>I35</b>	-0.25	-0.454	0.031	0.698	-0.025
<b>I18</b>	-0.017	-0.157	-0.07	-0.121	0.365	-	-	-	-	-	-

Source: prepared by the author (2023).

Analyzing the five categories, whose choosing probabilities are represented by parameters  $a_1$ ,  $a_2$ ,  $a_3$ ,  $a_4$  and  $a_5$  on Table 02, category  $a_3$ <sup>3</sup>, relative to the correct answer, is expected to present the highest value among the parameters. Furthermore, the NRM shows that the curve related to the correct answer is ascending in relation to the proficiency, that is, the hit probability increases when  $\theta$  - the respondent's latent trait, which can be understood as a measure of skill or proficiency - is higher (Pasquali, 2018)

However, there were questions in which the correct answer presented discrepant results in relation to their distractors, which could also be observed in the curves of some questions. Calculating  $\theta$  in NRM incorporates which wrong answers the students select, thus acknowledging that different wrong answers may indicate different levels of understanding (Smith; Bendjilali, 2022).

<sup>3</sup>  $a_s$  is the Nominal Response Model's category naming.

The next section refers to the question's quantitative analysis followed by their qualitative interpretation.

## 4.2 Quantitative and qualitative perspective for question evaluation

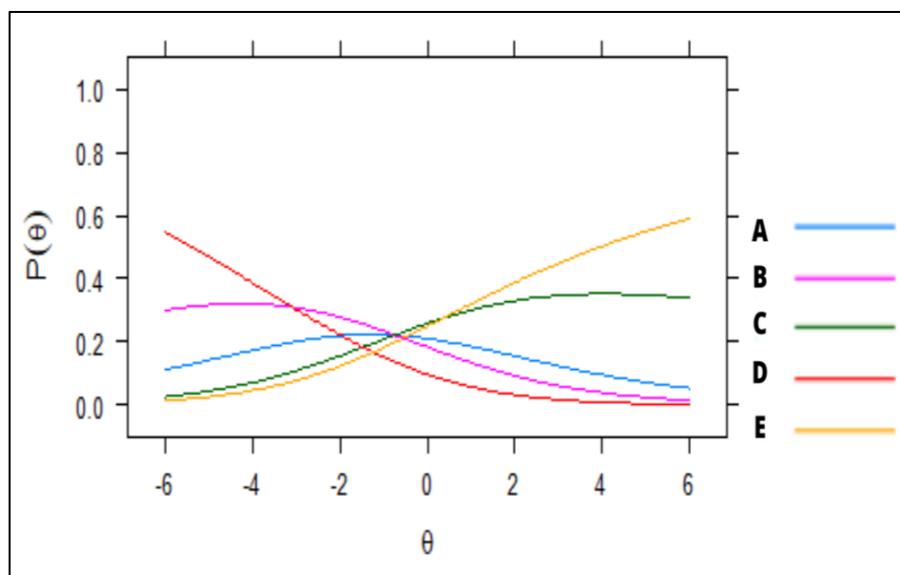
NRM-based question parameter estimation led to findings like those by Thissen, Steinberg and Fitzpatrick (1989), that is, questions presenting higher and positive parameters associated to the correct category ( $a_{correct}$ ), while distractor categories show lower or negative scores. With that in mind, we will detail questions I9, I16, I19, I21 and I23, which are characterized by distractors with parameters and curves unlike the model standard, and analyze them from the perspective of error analysis in mathematics.

Question 09 has answer C as parameter  $a_3=0.261$ , however distractor E presented a higher estimated probability ( $a_5=0.359$ ). In Figure 01<sup>4</sup>, we noticed that curve A stabilizes as  $\theta$  increases, staying under E. Notice that Table 01 shows equivalent approximate proportions between the correct answer and distractor E.

---

<sup>4</sup> The graphs representing the curves from the questions, found in Figures 01, 03, 05, 07 and 09, may present ascending or descending curves, related to a higher or lower probability of an alternative being chosen by the respondent. The correct alternative is usually chosen by individuals with greater proficiency. The skill/proficiency scale in the graph is organized from -6 to +6, which is R's presentation standard, allowing for a better visualization of the score amplitude.

Figure 01 - Curves of question 09



Source: prepared by the author (2023).

These data allow for a qualitative analysis of the question (Figure 02) that approaches integrals. This question requires the acknowledgment of the application of the definite integral as the area of regions delimited by functions, as well as the understanding that the numerical value of a negative function is negative. Consequently, in answer II, the result of the integral does not represent the area of the region described by the curves. The value 3 indicated in answer II acted as a distractor, leading respondents to the wrong alternative. For Cury (2013), without mastering the skills, such as the ability to deal with rules to calculate limits, derivatives or integrals, the student has no tools to work with the concepts. This seems to be harming the correct evaluation of question 09's response.

Figure 02 - Enade 2017's question 09 for the undergraduate program in Mathematics

**Question 09**

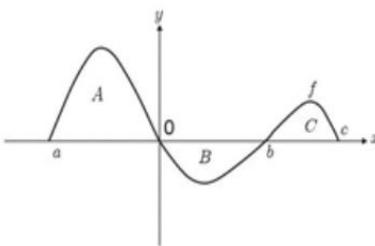
Let  $f: [a, c] \rightarrow \mathbb{R}$  be a continuous function and  $b \in (a, c)$ , as shown in the graph below.

$A$  is the area limited by  $y = 0$  and the curve  $\{(x, f(x)); x \in [a, 0]\}$ ;

$B$  is the area limited by  $y = 0$  and the curve  $\{(x, f(x)); x \in [0, b]\}$ ;

$C$  is the area limited by  $y = 0$  and the curve  $\{(x, f(x)); x \in [b, c]\}$ .

Knowing that  $A = 5$ ,  $B = 3$  and  $C = 2$ , consider the following statements.



I.  $\int_a^0 f(x)dx = 5$

II.  $\int_0^b f(x)dx = 3$

III.  $\int_a^c f(x)dx = 4$

The correct statements are:

A) I, only

B) II, only

C) I and III, only

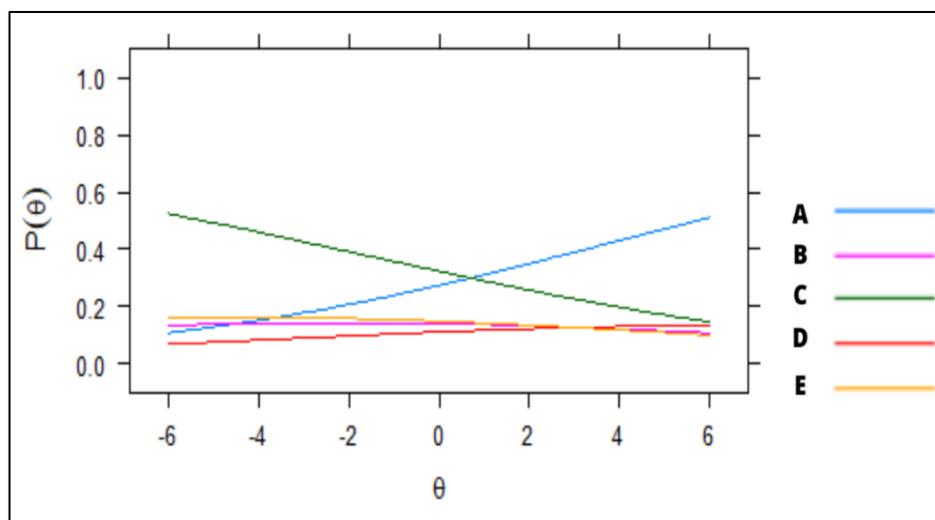
D) II and III, only

E) I, II and III

Source: Adapted from Enade (2017).

Estimates for question 16 show a probability of correct answer D of 0.052 ( $a_4$ ), while distractor A presents a probability of 0.125 ( $a_1$ ). Table 01 also informs that the choosing proportion of the correct answer was lower than A's (11.1 % and 27.4 %, respectively). However, distractor C was the most frequently chosen (32.6 %), which amplifies the distortions for this question. Graphically, figure 03 shows that curve A, relative to alternative A, increases with  $\theta$ , remaining with low probability for the correct answer (D).

Figure 03 - Curves from question 16



Source: prepared by the author (2023).

In question 16, the error stems from the respondents' substituting the values in the system variables, without analyzing if the problem has other solutions or relating to the mathematical foundations of linear algebra or analytical geometry. According to Cury (2013), this type of mistake indicates that respondents do not understand the process that must take place and try to deduce what should be done from the information they have.

Figure 04 - Enade 2017's question 16 for the undergraduate program in Mathematics

### Question 16

The solution of a linear system with three equations and three variables can be geometrically interpreted as the intersection of three planes in space and consists in checking if the three planes have only one point, infinite points or no points in common to determine if the system has only one solution, infinite solutions or no solution, respectively. Based on that information, it can be concluded that the linear system:

$$\begin{cases} x + 2y + z = 1 \\ y + z = 2 \\ x + y = -1 \end{cases}$$

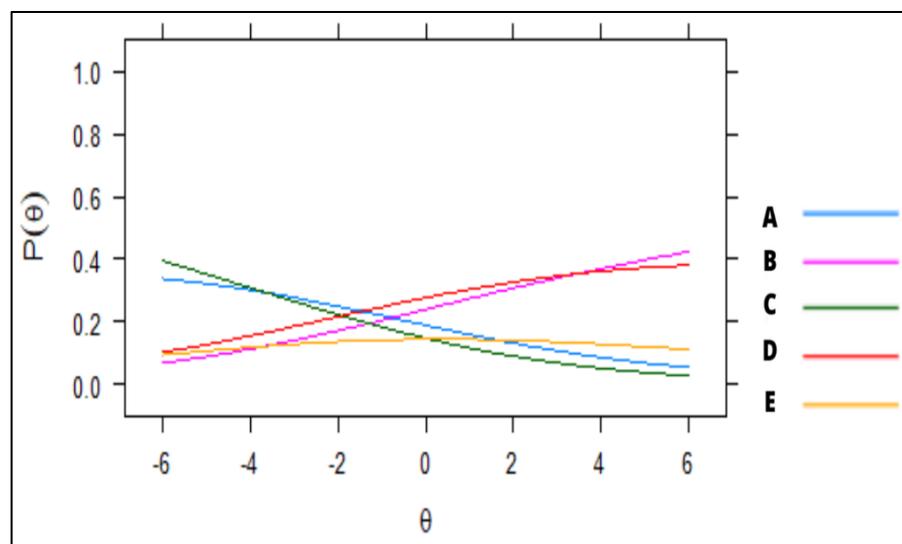
has, as a solution:

- A) the point (0, -1,3).
- B) the plane that goes through point (0, -1,3) and has the normal vector  $n=(1,2,1)$ .
- C) the line that goes through point (0, -1,3) and has the direction vector  $v=(1,2,-1)$ .
- D) the line that goes through point (0, -1,3) and has the direction vector  $v=(1,-1,1)$ .
- D) an empty set.

Source: Enade (2017).

Regarding question 19, there were two distractors with a positive probability (B with  $a_2=0.171$  and E with  $a_5=0.033$ ). The correct answer (D) had a probability of 0.129 ( $a_4$ ), thus being lower than distractor B. Figure 05 shows the curve relative to alternative D slowly decreasing in relation to B.

Figure 05 – Curves from question 19



Source: prepared by the author (2023).

In the question resolution (Figure 06), we can infer that the large number of steps may have led the respondents to misinterpret the first results: the given situation must be interpreted geometrically and, for that, they should switch between the numerical and graphical representations of the given points (which represented the cities). At this point, it's important to explore Duval (2012), who states that resorting to many registers seems to be a necessary condition for the mathematical objects not to be confused, and so that they can be recognized in each of their representations. The fragility in this transition between representations and in the knowledge of analytical geometry may have contributed to such results presented by the respondents. This aspect is also analyzed by Cury (2013), who talks about classes of mathematical errors in which the respondents can't properly conclude the thought development due to the lack of understanding of some substeps.

Figure 06 - Enade 2017's question 19 for the undergraduate program in Mathematics

### Question 19

According to the national policy for solid residues, implementing the collection of recyclable materials is a responsibility of the municipal administration, and recyclable collection goals must be part of a city's management plan.

Suppose the mayors of three neighbor cities have decided to work on a joint project to build a recyclable material selection warehouse, which could also function as a voluntary drop-in collection point. They have decided that the warehouse should be built at a location equidistant from the centers of the cities, represented, on the cartesian plane, by points  $A = (1,3)$ ,  $B = (10,0)$  and  $C = (3,7)$ .

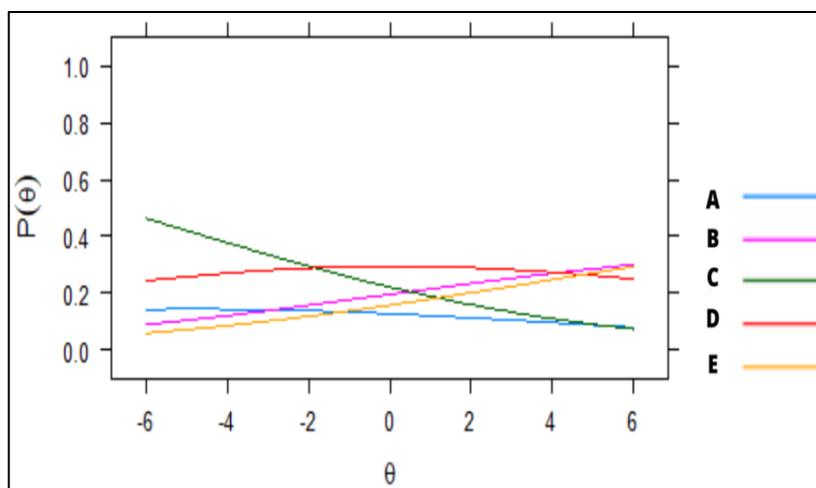
Under those circumstances, what are the chosen point's coordinates?

- A) (5,3)
- B)  $\left(\frac{11}{2}, \frac{7}{2}\right)$
- C)  $\left(6, \frac{5}{2}\right)$
- D) (6,3)
- E)  $\left(\frac{13}{2}, \frac{7}{2}\right)$

Source: Enade (2017).

Item 21 has distractors B and E with probabilities ( $a_2=0.094$  and  $a_5=0.126$ ), while the correct answer has a probability - 0.005 ( $a_4$ ). When analyzed through the CTT, this item presented negative point-biserials for all alternatives, with the highest value referring to the correct answer (-0.01), but very close to the distractors B and E (-0.07 and -0.05, respectively). Curve D (Figure 07), decreasing as  $\theta$  increases, confirms the presented quantitative incoherences.

Figure 07 - Curves from question 21



Source: prepared by the author (2023).

When solving I21 (Figure 08), some aspects must be analyzed regarding the sum of probabilities from answer I. The first calculations lead to three fractions equal to  $1/15$ . The need to add the values must be understood. Otherwise, alternative E is wrongfully selected. Such discrepancy can be observed in Table 06, which shows a proportion of 30 % of the participants choosing that distractor. In line with Viali and Cury (2009), this error can be considered a subclass of a situation in which the respondent understands the concept of probability as the ratio between the number of favorable cases and the number of possible cases, but has trouble discerning composite events.

Figure 08 - Enade 2017's question 21 for the undergraduate program in Mathematics

### Question 21

Six students signed up for a school chess championship: three girls, two of which are twin sisters, and three boys. For the first round, three sets of two competitors will be formed randomly as follows: the first player is drawn from the six participants, the second is drawn from the remaining five, the third from the remaining four, the fourth from the remaining three. The first pair of competitors is composed by the two players who were drawn first, the second by the two next ones, and the third is formed by the two players who were not drawn.

Considering those conditions for the selection of competitors, analyze the following statements:

- I. The probability of the twins competing against each other is  $1/15$ .
- II. The probability of the first pair of competitors being only boys is  $1/5$ .
- III. The probability of the first pair of competitors being one boy and one girl is  $3/5$ .

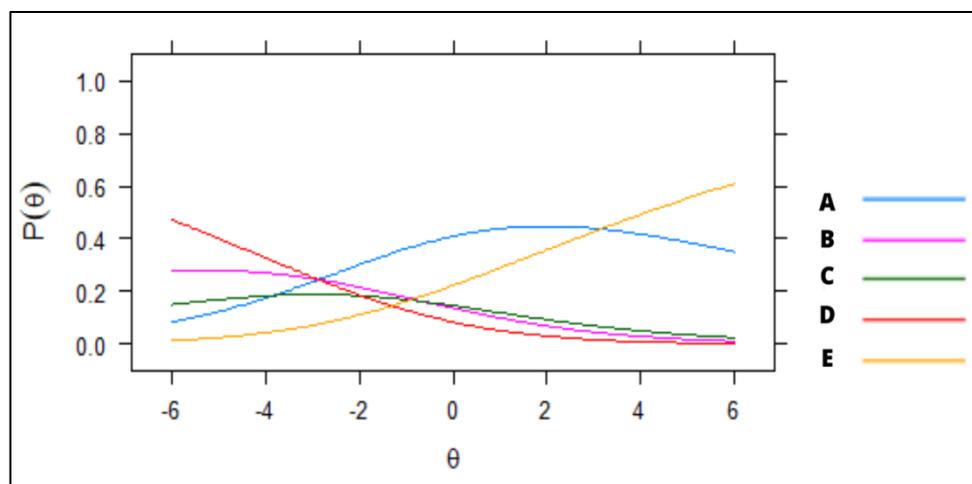
The correct statements are:

- A) I, only.
- B) II, only.
- C) I and III, only.
- D) II and III, only.
- E) I, II and III.

Source: Enade (2017).

For question 23 (Figure 09), a probability of 0.396 ( $a_5$ ) for the correct answer E was estimated through NRM, while distractor A had 0.202 ( $a_1$ ). Also regarding the choice of respondents, 20 % chose answer A and 40 % went for answer B. However, the CTT evaluation showed positive beserials for alternative A (0.01) and for the correct answer E (0.03). The graph model shows that curves B, C and D have reduced probability as the value of  $\theta$  advances. Curve A goes slightly up for some values of  $\theta$ , being overcome by the correct answer, however indicating the choice for the distractor by high-skilled respondents.

Figure 09 - Curves for question 23



Source: prepared by the author (2023).

Considering those aspects, the content of question 23's distractors (Figure 10) showed that respondents are drawn to distractor A, since the relevant calculations lead to the described value of  $-33/2$ . However, the function has a point of discontinuity indicating that 0 is part of the interval  $[-2,1]$ , establishing an undefined function and validating the correct answer E. Cury and Cassol (2004) talk about errors related to learning calculus and errors involving the number interval that is part of the problem situation.

Figure 10 - Enade 2017's question 23 for the undergraduate program in Mathematics

Question 23

Considering that a student is testing a piece of software to calculate the integral  $\int_{-2}^1 \left(\frac{1}{x^2} - 5\right) dx$ , analyze the following statements and the proposed relationship between them.

I. The result  $\int_{-2}^1 \left(\frac{1}{x^2} - 5\right) dx = -\frac{33}{2}$  given by the software is correct.

BECAUSE

II. The primitive of the function  $f(x) = \frac{1}{x^2} - 5$  is the function  $F(x) = -\frac{1}{x} - 5x$  and, through the fundamental theorem of calculus, it can be concluded that

$$\int_{-2}^1 \left(\frac{1}{x^2} - 5\right) dx = \left(-\frac{1}{x} - 5x\right) \Big|_{-2}^1 = \left(-\frac{1}{1} - 5 \cdot 1\right) - \left(-\frac{1}{(-2)} - 5(-2)\right) = -\frac{33}{2}$$

Regarding the statements above, choose the correct option:

A) Statements I and II are true, and II is a correct explanation for I.  
 B) Statements I and II are true, but II is not a correct explanation for I.  
 C) Statement I is true, and II is false.  
 D) Statement I is false, and II is true.  
 E) Statements I and II are false.

Source: Adapted from Enade (2017).

Thus, within the perspectives of the 35 analyzed items, 5 (14 %) presented weakly designed alternatives (distractors). Questions I16, I19, I21 and I23 were disregarded for the score calculation by the systematic adopted by Enade 2017 (Brasil, 2018), since they presented too low of a discrimination index (biserial point correlation). However, they were considered in this study, since they were effectively answered by Enade 2017's participants. Therefore, understanding these items is technically necessary.

With respect to this withdrawal of questions by Enade 2017, the Reference Matrix (Brasil, 2018), with questions I16 and I19, proposed to assess the ability to "develop conjectures and generalizations by establishing relationships between formal and intuitive aspects", with only one item remaining with such expectation. As for question I21, the assessed skill was "problem solving". However, its removal weakened the test, since only two other questions (I20 and I21) were meant to assess the Probability and Statistics object of knowledge. In addition, question I23 is the only one

that brings elements of the object of knowledge “fundamentals of analysis”, creating an important evaluative gap in this regard. As for I9, a reevaluation by specialists during the development stage would be admissible, since it presents problems in its distractors.

In this analysis, the data show that the system adopted by INEP for Enade interferes with the assessment of the student’s competences in mathematics. Thus, Alves (2020) indicates that removing questions should not be the starting point, indicating that specialists should be consulted in order to assess if they are essential to validate the contents and if the recommendation for their withdrawal would be the most plausible and unbiased path.

## 5 Final considerations

This study intended to interpret the contents and alternatives of Enade 2017’s questions regarding the undergraduate program in mathematics, in order to contribute to the deepening of its edumetric analysis. In this sense, there was no specific interest to work on an IRT parameterization scenario, using the 3-parameter logistic model, but to obtain data that would allow to evaluate not only the technical aspects of item formulation, but also the possible reasoning that implies the understanding of gaps in the training of these prospective teachers, which is crucial for didactic-pedagogical actions in programs within the sphere of Mathematics Education.

In this context, the test analysis went from the general to the specific, identifying possible discrepancies in the behavior of the parameters of each of the alternatives. The parameters were expected to reach higher values as the latent trait progressed, a situation that could be observed in the graphs of the question curves. However, the proportion of 5/27 of these items differed from the expected approach, indicating that the Enade 2017 test had questions whose distractors should have been better analyzed in terms of their technical aspect. Furthermore, they reveal weaknesses in students’

mathematical reasoning, suggesting difficulties in the semiotic understanding associated with mathematical language and in calculus and algebra techniques.

The application of the Nominal Response Model (NRM) enables to create graphical representations that point out more attractive options for students. In this model, the most frequently chosen answers by the students, adapted to their proficiency levels, can be identified. This enables a comprehensive analysis that considers both the totality and individual nuances. Within this approach, it is also feasible to estimate the discrimination parameters, which set apart the values associated with the most likely responses, as discussed by Reise *et al.* (2023). The NRM provides a consistent foundation for an analysis that combines quantitative and qualitative elements, allowing detailed observations about the evolution of students and their skills. Besides, this model can explain the choice of less correct alternatives, even revealing random guesses. This ability is extremely important to identify critical points and gaps in the knowledge of the respondent group. Thus, it enables a supplementary analysis of a qualitative nature, by tracking possible student mistakes and connecting them to investigations related to error analysis.

The results also indicate the need to question administrative entities regarding the experimentation of new systems that can provide greater precision to the results. These, in turn, imply the reception, by society, of an educator with potential knowledge for their professional experiences, which expands the opportunity to offer quality education to citizens.

In this context, the approaches by Vianna (2009), who argues that the results of educational assessments should not be used solely and exclusively to translate a certain performance, should be emphasized. They can involve the definition of new public policies, of projects for the implementation and modification of curricula, of continuous training programs for teachers and, crucially, of elements for decision-making that aim to have an impact, that is, to change the thinking and actions of members of the

system. Thus, the results obtained via NRM have the potential to not only improve the technical construction of the test, but also impact changes in the Enade assessment guidelines and in the curricular structure of initial training courses.

## Acknowledgements

We would like to thank Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina - FAPESC (grant No. 2023TR000334); the State of Santa Catarina College Scholarship Program - UNIEDU; and the research group NEPESTEEM.

## References

- AGRESTI, A.; KATERI, M. Categorical data analysis. *In: LOVRIC, M. (ed.) International encyclopedia of statistical science*. Heidelberg: Springer, 2011. p. 206–208. Disponível em: [https://link.springer.com/referenceworkentry/10.1007/978-3-642-04898-2\\_161](https://link.springer.com/referenceworkentry/10.1007/978-3-642-04898-2_161). Acesso em 20 out. 2023.
- ALLEVATO, N. S. G. Resolução de problemas, software gráfico e detecção de lacunas no conhecimento da linguagem algébrica. *In: ENCONTRO NACIONAL DE EDUCAÇÃO MATEMÁTICA - ENEM, 8., 2004, Recife. Anais [...]*. Recife: SBEM, 2004. p. 1-20. Disponível em: <http://www.sbembrasil.org.br/files/viii/pdf/06/CC47973757953.pdf>. Acesso em: 07 out. 2023.
- ALVES, C, K. T. **Análise das propriedades psicométricas da prova de conhecimentos específicos de licenciatura em ciências biológicas no ENADE 2017**. 2020. Dissertação (Mestrado Profissional em Métodos e Gestão em Avaliação) - Universidade Federal de Santa Catarina, Florianópolis, 2020. Disponível em: <https://repositorio.ufsc.br/bitstream/handle/123456789/216664/PMGA0048-D.pdf?sequence=-1&isAllowed=y>. Acesso em: 16 maio 2023.

BELTRÃO, K. I.; MANDARINO, M. C. F. Análise dos itens de múltipla escolha das provas do Enade 2016. **Estudos em Avaliação Educacional**, São Paulo, v. 34, p. e07951, 2023. Disponível em:

<https://publicacoes.fcc.org.br/eae/article/view/7951>. Acesso em: 29 ago. 2023.

BOCK, R. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. **Psychometrika**, Heidelberg, v. 37, n. 1, p. 29-51, 1972. Disponível em: <https://link.springer.com.ez74.periodicos.capes.gov.br/article/10.1007/BF02291411>. Acesso em: 07 out. 2023.

BRASIL. Instituto Nacional de Pesquisas e Estudos Educacionais Anísio Teixeira. **Relatório síntese de área: matemática**. Brasília: Inep, 2017. Disponível em: [https://download.inep.gov.br/educacao\\_superior/enade/relatorio\\_sintese/2017/Matematica.pdf](https://download.inep.gov.br/educacao_superior/enade/relatorio_sintese/2017/Matematica.pdf). Acesso em: 16 maio 2023.

CAZORLA, I. M. **A relação entre a habilidade viso-pictórica e o domínio de conceitos estatísticos na leitura de gráficos**. 2002. Tese (Doutorado em Educação) - Universidade de Campinas, Campinas, 2002. Disponível em: [https://www.psiem.fe.unicamp.br/pf-psiem/cazorla\\_irenemauroicio\\_d.pdf](https://www.psiem.fe.unicamp.br/pf-psiem/cazorla_irenemauroicio_d.pdf). Acesso em: 16 maio 2023.

CHALMERS, R. P. Generating adaptive and non-adaptive test interfaces for multidimensional item response theory applications. **Journal of Statistical Software**, EUA, v. 71, p. 1-38, 2016. Disponível em: <https://www.jstatsoft.org/article/view/v071i05>. Acesso em: 07 out. 2023.

CURY, H. N. **Análise de erros: o que podemos aprender com as respostas dos alunos**. Belo Horizonte: Autêntica, 2013.

CURY, H. N.; CASSOL, M. Análise de erros em cálculo: uma pesquisa para embasar mudanças. **Acta Scientiae: revista de ensino de ciências e matemática**, São Paulo, v. 6, n. 1, p. 27-36, 2004. Disponível em: <http://www.periodicos.ulbra.br/index.php/acta/article/view/128/116>. Acesso em: 07 out. 2023.

DUVAL, R. THADEU, M. (tradução). Registros de representação semiótica e funcionamento cognitivo do pensamento. **Revemat: revista eletrônica de matemática**, Santa Catarina, v. 7, n. 2, p. 266-297, 2012. Disponível em: <https://periodicos.ufsc.br/index.php/revemat/article/view/1981-1322.2012v7n2p266>. Acesso em: 07 out. 2023.

HALADYNA, T. M.; DOWNING, S. M.; RODRIGUEZ, M. C. A review of multiple-choice item-writing guidelines for classroom assessment. **Applied Measurement in Education**, São Paulo, v. 15, n. 3, p. 309-333, 2002. Disponível em: <http://site.ufvjm.edu.br/fammuc/files/2016/05/item-writing-guidelines.pdf>. Acesso 07 out. 2023.

HOSMER, D. W.; LEMESHOW, S. **Applied logistic regression**. Wiley: New York, 2000.

KRUTETSKII, V. A. **The psychology of mathematical abilities in schoolchildren**. Chicago: University of Chicago Press, 1976.

KUTNER, M. H.; NACHTSHEIM, C. J.; NETTER, J.; LI, W. **Applied linear statistical models**. Boston: McGraw-Hill Irwin, 2005. Disponível em: [https://users.stat.ufl.edu/~winner/sta4211/ALSM\\_5Ed\\_Kutner.pdf](https://users.stat.ufl.edu/~winner/sta4211/ALSM_5Ed_Kutner.pdf). Acesso em: 07 out.

LOPES, F. L.; VENDRAMINI, C. M. M. Propriedades psicométricas das provas de pedagogia do Enade via TRI. **Avaliação: revista da avaliação da educação superior**, Campinas; Sorocaba, v. 20, n. 1, p. 27–47, mar. 2015. Disponível em: <https://www.scielo.br/j/aval/a/64h6cvwZrKPR9v4P4nChZ6t/?format=pdf&lang=pt>. Acesso em: 07 out. 2023.

MARTINS, V. B. **Habilidades de estudantes do ensino médio para planificar figuras tridimensionais**. 2018. Monografia (Especialização em Ciências e Tecnologias da Educação) - Instituto Federal Sul – Riograndense, Pelotas, 2018. Disponível em: <https://viniciuscavq.files.wordpress.com/2018/06/2018-vanderlei-planificac3a7c3a3o.pdf>. Acesso em: 16 maio 2023.

PASQUALI, L. **TRI – teoria da resposta ao item: teoria, procedimentos e aplicações**. Curitiba: Apris, 2018.

PEREIRA, A.; OLIVEIRA, I.; TINOCA, L. A cultura de avaliação: que dimensões. *In*: ENCONTRO INTERNACIONAL TIC E EDUCAÇÃO: TICeduca, 1., 2010, Lisboa. **Actas** [...]. Lisboa, 2010. p. 350-357.

PINHEIRO, I. R.; COSTA, F. R.; CRUZ, R. M. Modelo nominal da teoria de resposta ao item: uma alternativa. **Avaliação Psicológica**, São Paulo, v. 9, n. 3, p. 437-447, 2010. Disponível em: [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1677-04712010000300010](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712010000300010). Acesso em: 07 out. 2023.

R CORE TEAM. **R: a language and environment for statistical computing**. R Foundation for Statistical Computing: Vienna, Áustria, 2022. Disponível em: <https://www.R-project.org/>. Acesso em: 17 maio 2023.

REISE, S. P.; HUBBARD, A. S.; WONG, E. F.; SCHALET, B. D.; HAVILAND, M. G.; KIMERLING, R. Response category functioning on the health care engagement measure using the nominal response model. **Assessment**, [S. l.], v. 30, n. 2, p. 375-389, 2023. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10262955/>. Acesso em: 07 out. 2023.

REVELLE, M. W. Procedures for psychological, psychometric, and personality. **R Package Biopsy**, Northwestern University, Evanston, Illinois. R package version 2.3.6, 2023. Disponível em: URL: <https://CRAN.R-project.org/package=psych>. Acesso em: 07 out. 2023.

SILVA, A. F. Z. da; ANDRADE, D. F. de; BORGATTO, A. F.; NAKAMURA, L. R. Aplicação do modelo de reposta nominal da TRI a avaliação educacional de larga escala. **Sigmae**, Minas Gerais, v. 8, n. 2, p. 735-741, 2019. Disponível em: <http://publicacoes.unifal-mg.edu.br/revistas/index.php/sigmae/article/view/1036/693>. Acesso em: 07 out. 2023.

SMITH, T. I.; BENDJILALI, N. Motivations for using the item response theory nominal response model to rank responses to multiple-choice items. **Physical Review Physics Education Research**, [S. l.], v. 18, n. 1, p. 1-13, 2022. Disponível em: <https://journals.aps.org/prper/pdf/10.1103/PhysRevPhysEducRes.18.010133>. Acesso em: 07 out. 2023.

STEWART, J.; DRURY, B.; WELLS, J.; ADAIR, A.; HENDERSON, R.; MA, Y.; LEMONCHE, A. P.; PRITCHARD, D. Examining the relation of correct knowledge and misconceptions using the nominal response model. **Physical Review Physics Education Research**, [S. l.], v. 17, n. 1, p. 1-15, 2021. Disponível em: <https://journals.aps.org/prper/pdf/10.1103/PhysRevPhysEducRes.17.010122>. Acesso em: 07 out. 2023.

THISSEN, D.; STEINBERG, L.; FITZPATRICK, A. R. Multiple-Choice Models: the distractors are also part of the item. **Journal of Educational Measurement**, EUA, v. 26, n. 2, p. 161-76, 1989. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/j.1745-3984.1989.tb00326.x>. Acesso em: 20 out. 2023.

VIALI, L.; CURY, H. N. Análise de erros em probabilidade: uma pesquisa com professores em formação continuada. **Educação Matemática Pesquisa**, [S. l.], v. 11, n. 2, p. 373-391, 2009. Disponível em: <http://funes.uniandes.edu.co/24348/1/Viali2009An%C3%A1lise.pdf>. Acesso em: 07 out. 2023.

VIANNA, H. M. Fundamentos de um programa de avaliação educacional. **Revista Meta: Avaliação**, Rio de Janeiro, v. 1, n. 1, p. 11-27, abr. 2009. Disponível em: <https://revistas.cesgranrio.org.br/index.php/metaavaliacao/article/view/11/4>. Acesso em: 29 ago. 2023.

ZHANG, X.; ZHAO, C.; XU, Y.; LIU, S.; WU, Z. Kernel causality among teacher self-efficacy, job satisfaction, school climate, and workplace well-being and stress. **TALIS. Frontiers in Psychology**, EUA, v. 12, p. 1-16, 2021. Disponível em: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.694961/full>. Acesso em: 07 out. 2023.

## **With contribution from the authors**

Sandra Cristina Martini Rostirola - Data collection and analysis and text development.

Elisa Henning - Result analysis, advising and text development.

Ivanete Zuchi Siple - Result analysis, advising and text development.

**Portuguese language review:** Marcia Vidal.

E-mail: frozzamarciavidal@gmail.com

**Translation to English:** Diogo Guedes de Figueiredo

E-mail: diogoguedesdefigueiredo@gmail.com