# EVIDENCE ON THE INTERNAL STRUCTURE OF ENEM AND THE COMPETENCIES OF NATURAL SCIENCES[1]

**Rodrigo Travitzki[2] and Ricardo Primi[3]**

**ABSTRACT**

In this work, the internal structure of ENEM is investigated, with special attention to the Natural Sciences test. Microdata from 2014 and 2015 were analyzed. Factor analysis suggests that one factor is sufficient to describe the variance of the 175 items. The two-factor model suggests systematic heterogeneity, in the two years, between the sub-factors, especially in the Mathematics and Languages and Codes tests. On the other hand, the average of the 4 tests proved to be collinear with the general factor. The Natural Sciences test showed considerable internal multidimensionality, which could be a characteristic of scientific literacy. The Natural Sciences competencies described in the Reference Matrix were consistent with the empirical behavior of the items, suggesting that the internal structure of the test reflects its theoretical foundations.

**Keywords:** ENEM. Scientific literacy. Educational evaluation. Validity. Two-factor model.

[2] Highest degree: PhD in Education from the University of São Paulo
Academic institution: Unicamp, Campinas
Email: r.travitzki@gmail.com
ORCID: https://orcid.org/0000-0002-0107-682X
[3] Highest degree: PhD in School Psychology and Human Development from the University of São Paulo
Academic institution: University of São Francisco, Campinas
Email: ricardo.primi@usf.edu.br
ORCID: https://orcid.org/0000-0003-4227-6745

## INTRODUCTION

The National High School Exam (ENEM) is an important reference for Brazilian education and there is a need for studies on the validity of the tests from 2009, when the Exam underwent structural transformations. There is evidence that the first editions of the ENEM evaluated reasoning capacity (fluid intelligence) more than knowledge itself (crystallized intelligence) (PRIMI et al., 2001). Regarding recent editions, there are few validity studies (GOMES, GOLINO, SOUZA PERES, 2020; TRAVITZKI, 2017).

The internal structure is one of the sources of information to verify the validity of a test (AERA; APA; NCME, 2014). In the case of ENEM, it is necessary to investigate to what extent the items of the four tests correspond to four minimally distinct constructs. It is also necessary to verify whether there is a construct underlying the items of the Natural Sciences test and, if so, whether this construct can be interpreted as a form of scientific reasoning. Such objectives guide this work.

In relation to the Natural Sciences, there is evidence that children aged 4 to 6 years have basic elements of scientific reasoning, such as the relationship between hypothesis and evidence, or even perception of covariance (KOERBER et al., 2005). According to Stephen Norris and Linda Phillips, scientific literacy has a particularity in relation to other types of literacy, or literacy,[4] because it depends more on specific previous knowledge. In this sense, the authors consider that there are two aspects of scientific literacy: 1) the fundamental (ability to understand, interpret, analyze and criticize any text) and 2) the derivative (dependent on specific content of each science). However, they conclude, the fundamental meaning is little applied in science education (NORRIS; PHILLIPS, 2003), which may be related to the relatively greater importance of specific content in this type of literacy, which is sometimes overvalued.

This article aims to contribute to the understanding of the internal structure of ENEM, applying psychometric techniques in the exam as a whole and also particularly in the Natural Sciences test. The first results refer to the internal structure in general terms, without taking into account the ENEM Reference Matrix, focusing mainly on the two-factor model. Next, results are presented regarding the coherence between the Natural Sciences competencies ofthe Reference Matrix and the empirical behavior of the items.

---

[4] The original term in English is *literacy*, which is translated in Brazilian literature sometimes as "literacy", sometimes as "literacy" (SASSERON, CARVALHO, 2011), which in this work are considered synonyms.

**METHODOLOGY**

Initially, a factor analysis of the 175 items[5] of each year was performed, based on tetrachoric correlations. Factor analysis is one of the psychometric foundations for the investigation of intelligence (PRIMI et al., 2001). In the context of educational evaluations, it is common to observe a predominant factor in the test. An important question, in this case, is to verify to what extent this predominant factor corresponds to the planned construct, and to what extent it corresponds to more generic skills, related to the resolution of tests, such as fluid intelligence and processing speed – in the nomenclature of the Cattell-Horn-Carroll Theory (PRIMI, 2003).

To seek answers to this question, two-factor models were also used, which were initially proposed as an alternative to Thurstone's simple oblique structure, contemplating different levels of complexity in psychological behavior (SCHMID; LEIMAN, 1957). In fact, this type of model allows the predominant factor of the test to be isolated in order to investigate more specific aspects of its internal structure. A general factor (present in all items) and some subfactors (present in subsets of items) are estimated, which would correspond to conceptually more specific constructs. It is usually assumed that the general factor is orthogonal to the sub-factors. Two-factor models allow: a) to investigate the partitioning of variance when it is believed that there is a more general factor and some sub-factors; b) the control of multidimensionality in essentially one-dimensional tests; c) to evaluate whether the general factor is strong enough to justify one-dimensional models; d) determine the adequacy of the overall score, and whether there is any gain with the inclusion of sub scores (RODRIGUEZ; REISE; HAVILAND, 2016).

For the analysis of the coherence between NC competencies and the empirical behavior of the items, the inter-item tetrachoric correlations of all 45 items were calculated. After that, the averages of these correlations were calculated, grouped by competency. The same procedure was repeated, grouping the correlations by discipline, for comparison purposes.

The 2014 and 2015 ENEM microdata were analyzed with the R software, Psych package (REVELLE, 2018). The sample included only high school graduates, from regular schools, present on both days. All notebooks were included, with one exception.[6] To this end, the items were reordered with the information provided in the microdata themselves. The reference for the ordering of items was the blue notebook for Saturday – Natural

---

[5] Considering that 5 items of the Languages and Codes test are foreign languages and, therefore, will be removed from the analysis.

[6] The pink notebook of the 2014 HC test was removed because it presented an inconsistency, probably related to the table that describes the correspondence between items and notebooks.

Sciences (NC) and Human Sciences (CH) – and the yellow notebook for Sunday – Mathematics (MT) and Languages and Codes (LC). A random sample of 300 thousand students was used, due to computational limitations.

## INTERNAL STRUCTURE OF ENEM

The eight tests presented an acceptable Cronbach's alpha coefficient (greater than 0.7), with the lowest values in the NC tests (0.76 in 2014 and 0.79 in 2015). Analyzing the set of four tests, the alpha coefficient was 0.95, although some items were found to be negatively correlated with the whole (4 in 2014 and 3 in 2015). It should be noted that an alpha coefficient greater than 0.9 may indicate an excessive number of items.

The parallel analysis of the 175 items revealed, in the two years, a predominant factor and a second minor factor that was also prominent, in addition to several other potentially significant ones. Analyzing the four tests separately, the MT and NC tests had a higher number of components than the other two (especially NC, with 7 components in 2014 and 9 in 2015).[7] Such results are compatible with the hypothesis of Norris and Phillips (2003), related to the greater need for prior specific knowledge for scientific literacy. The factor analysis revealed that one factor would be enough to contemplate the common variance. On the other hand, a more complete analysis can be performed with three factors, considering that the one-factor analysis captured 12% of the total variance, while the three-factor analysis captured 14%.

A two-factor model with three sub-factors was also estimated for each year, covering the 175 items. The hierarchical omega coefficient estimates the proportion of variance in the total score that can be attributed to the general factor, treating the "resulting" variance of the subfactors as measurement error (RODRIGUEZ; REISE; HAVILAND, 2016). When compared to total omega, it indicates the adequacy of one-dimensional models and possible gains with multidimensional models. In ENEM 2014, the hierarchical omega coefficient was 0.77 and in 2015 it was 0.72, and in the two years the total omega coefficient was 0.95. In fact, about one-fifth of the variance in single scores can be attributed to multidimensionality.

However, when the average in ENEM is observed, the impact of multidimensionality is irrelevant. The correlation between the mean of the 4 grades and the overall factor score in the two-factor model is 0.97 in 2015. Regarding the sub-factors, it does not reach 0.4. The correlation between the overall factor score and the 4 scores separately is slightly lower, especially in NC and MT (0.8), but it is still high.

---

[7] Number of components with eigenvalue greater than random eigenvalue in parallel analysis.

The Natural Sciences score, still in 2015, showed a higher correlation with the general factor (0.80), followed by the F2 sub-factor (0.33) – which is also correlated with MT (0.28). Such results suggest little specificity associated with the NC score, making it difficult to identify a clear construct being evaluated by the ENEM Natural Sciences test.

To analyze the relationship between the structure identified in the two-factor model and the structure of four ENEM tests, we calculated the coefficient of congruence between the factors of the model and the tests (Table 1). Such a coefficient can be interpreted as a correlation. A similar internal structure is noted in both years, especially clear in the TM and LC tests. Taking into account that the sub-factors include the residuals of the general factor, this means that, in addition to the general skills necessary for the resolution of tests, there are clear differences between the skills and competencies evaluated by the MT and LC tests. Although such differences have little impact on ENEM scores, as calculated today.

**Table 1: Coefficients of congruence**

|  | 2014 | | | | 2015 | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | CH | CN | LC | MT | CH | CN | LC | MT |
| g | 0,53 | 0,33 | 0,52 | 0,37 | 0,54 | 0,37 | 0,47 | 0,36 |
| F1 | 0,51 | 0,13 | **0,64** | **-0,02** | 0,52 | 0,09 | **0,63** | **-0,06** |
| F2 | 0,23 | 0,40 | **0,04** | **0,59** | 0,27 | 0,50 | **0,11** | **0,47** |
| F3 | 0,07 | 0,23 | **-0,09** | **0,63** | 0,22 | 0,25 | **-0,02** | **0,59** |
| h2 | 0,45 | 0,27 | 0,47 | 0,44 | 0,47 | 0,32 | 0,48 | 0,35 |

Note: coefficients of congruence between factors of the observed internal structure (in the two-factor model) and the ideally expected structure (four tests). The general factor (g), the three subfactors (F1 to F3) and the commonality (h2) are observed. Source: prepared by the authors with ENEM data.

In relation to the Natural Sciences test, it is again observed a certain proximity to the F2 sub-factor. However, this factor is also present in TM tests, making it difficult to identify a specific factor for NC. The same occurs between Human Sciences and F1.

The factor analysis of the 2015 Natural Sciences test suggests that a single factor would be sufficient to represent the common variation, explaining 10% of the variance in the 45 items. The parallel analysis shows the existence of a second prominent factor and a total of 15 factors (or 9 components) with a higher than expected self-value randomly. The alpha coefficient was 0.8 (considered adequate, assuming unidimensionality). In the two-factor model, the hierarchical omega was 0.64 (total omega = 0.81) and in 2014 it was 0.32 (0.78), suggesting that a single score does not adequately represent the internal structure of the NC test.

**EMPIRICAL EVIDENCE OF NATURAL SCIENCE COMPETENCIES**

We analyzed the 2015 NC test more deeply, with the objective of verifying to what extent the theoretical structure of skills and competencies is coherent with the empirical behavior of the items. It is expected that the correct answers in the items of a certain competence are more correlated with each other than with the items of the other competencies, which would be evidence of discrimination at the level of competencies. The ENEM Reference Matrix (INEP, 2009) describes 30 skills for the NC area, structured in the following 8 competencies:

1. Understand the natural sciences and the technologies associated with them as human constructions, perceiving their roles in the production processes and in the economic and social development of humanity.

2. Identify the presence and apply technologies associated with the natural sciences in different contexts.

3. Associate interventions that result in environmental degradation or conservation with productive and social processes and with scientific-technological instruments or actions.

4. Understand interactions between organisms and the environment, in particular those related to human health, relating scientific knowledge, cultural aspects and individual characteristics.

5. Understand methods and procedures specific to the natural sciences and apply them in different contexts.

6. To appropriate knowledge of physics in order to interpret, evaluate or plan scientific-technological interventions in problem situations.

7. To appropriate knowledge of chemistry in order to interpret, evaluate or plan scientific-technological interventions in problem situations.

8. To appropriate knowledge of biology in order to interpret, evaluate or plan scientific-technological interventions in problem situations.

The 2015 NC test presents at least one item for each of the 30 skills in the reference matrix. Of course, as there are 45 items in total, it is not possible to have 3 items for each skill, the minimum number recommended for measuring constructs in general. On the other hand, this condition is satisfied at the level of competencies, as can be seen in Figure 1, which informs the number of items for each competency.

**Figure 1: Average inter-item correlations among the 8 Natural Sciences competencies**
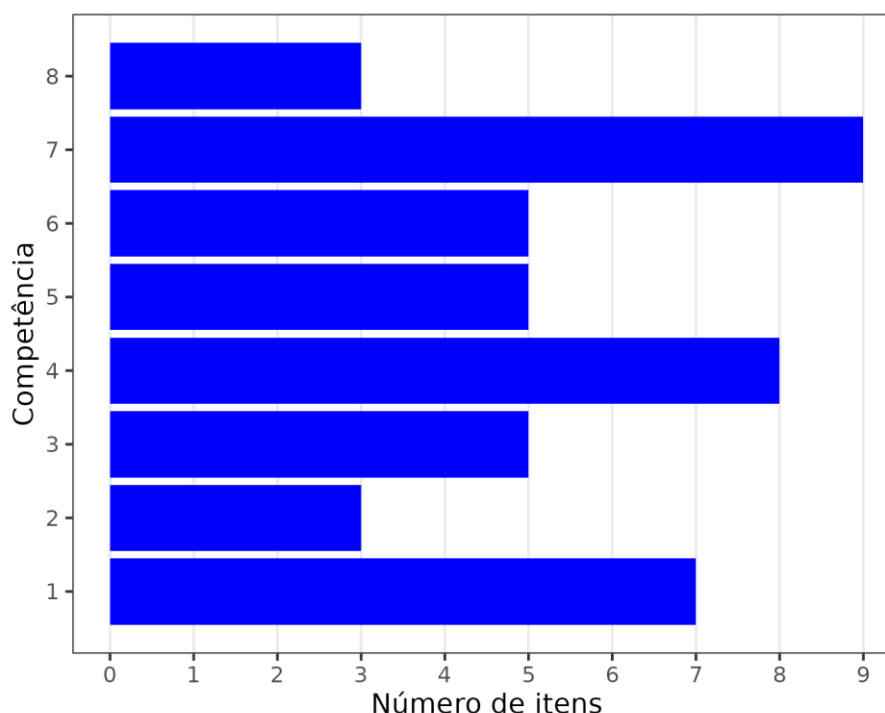


Source: prepared by the authors with data from ENEM 2015.

Figure 1 shows a certain imbalance between the competencies, i.e., there is considerable variation in the number of items relative to each one, which ends up overvaluing some competencies to the detriment of others in the calculation of NC proficiency. It is also worth highlighting, in this sense, competencies 6 to 8, which are specific to each discipline: there are 3 times more chemistry items (competency 7) than biology (competency 8). However, this is not necessarily a problem with regard to the balance between the three NC disciplines, as there are biology items in the other competencies, for example. According to our reading of the content of the test, there are 16 items with physics content, 13 chemistry and 15 biology, which can be considered sufficiently balanced. But with regard to the 8 competencies of the matrix, the lack of balance is clear.

To verify the coherence between the theoretical matrix and the empirical results, we calculated the inter-item correlation between all 45 test items, and then calculated the average of these correlations by grouping them by competency. For example, as shown in Table 2, the mean inter-item correlations between the 7 items of competency 1 is 0.19 while the mean inter-item correlations between these 7 items and the 3 items of competency 2 is 0.05. For a general reference, the average of all inter-item correlations of the 45 items of the NC test is 0.11.

**Table 2: Average inter-item correlations among the 8 Natural Sciences competencies**

|          | Comp. 1 | Comp. 2 | Comp. 3 | Comp. 4 | Comp. 5 | Comp. 6 | Comp. 7 | Comp. 8 |
|----------|---------|---------|---------|---------|---------|---------|---------|---------|
| Comp. 1  | **0,19** | 0,05   | 0,07    | 0,08    | 0,07    | 0,07    | 0,05    | 0,05    |
| Comp. 2  | 0,05    | **0,36** | 0,06    | 0,07    | 0,07    | 0,07    | 0,04    | 0,04    |
| Comp. 3  | 0,07    | 0,06    | **0,28** | 0,12    | 0,11    | 0,09    | 0,07    | 0,07    |
| Comp. 4  | 0,08    | 0,07    | 0,12    | **0,23** | 0,12    | 0,09    | 0,08    | 0,08    |
| Comp. 5  | 0,07    | 0,07    | 0,11    | 0,12    | **0,3** | 0,09    | 0,08    | 0,07    |
| Comp. 6  | 0,07    | 0,07    | 0,09    | 0,09    | 0,09    | **0,25** | 0,06    | 0,05    |
| Comp. 7  | 0,05    | 0,04    | 0,07    | 0,08    | 0,08    | 0,06    | **0,15** | 0,04    |
| Comp. 8  | 0,05    | 0,04    | 0,07    | 0,08    | 0,07    | 0,05    | 0,04    | **0,37** |

Source: prepared by the authors with data from ENEM 2015.

The most important information provided by Table 2 is that, in each of the 8 competencies, the highest inter-item correlations are those between the items of the competency itself (emphasis in bold). This result is evidence of the validity of the ENEM reference matrix, because if a respondent gets one of the items of a certain competency right, there is a greater probability of him getting the other items of that competency right, in relation to the items of other skills. In other words, the theoretical structure represented by the 8 competencies is consistent with the empirical behavior of the items.

In addition, it is worth noting that the competencies with the highest internal consistency (higher average of inter-item correlations) are 2 and 8. As can be seen in Figure 1, these are precisely the competencies with the lowest number of items. Similarly, the competencies with the least internal consistency are 1 and 7, which are among the competencies with the highest number of items. In fact, there seems to be a relationship between the number of items and the internal consistency of competencies, which is not unexpected.

Another point to highlight in Table 2 is that the inter-item correlations suggest a certain similarity between competencies 3, 4 and 5. A simple reading of the descriptors of these competencies does not reveal clear similarities between the three, although competencies 3 and 4 deal with the environmental theme to some extent. Therefore, such a result would need to be better understood, if confirmed in other editions of ENEM.

Finally, we performed a procedure similar to that of Table 2 for the three disciplines of Natural Sciences, resulting in Table 3. The information in Table 3 does not have the function of validating the NC test, as the reference matrix is structured in competencies and skills, not in disciplines. Therefore, the information in Table 3 can serve more as a complement or a reference for comparison, in order to better understand what is being evaluated by the NC items.

**Table 3: Mean inter-item correlations between the 3 disciplines of Natural Sciences**

|           | Physics | Chemistry | Biology |
|-----------|---------|-----------|---------|
| Physics   | **0,12**| 0,06      | 0,08    |
| Chemistry | 0,06    | **0,14**  | 0,08    |
| Biology   | 0,08    | 0,08      | **0,17**|

It is also noted that, also in Table 3, there is evidence of discrimination, as the correlations are greater between the items of the same discipline (emphasis in bold). On the other hand, the magnitude of the correlations is, as a rule, lower than that found when the items are separated by competence (Table 2). In fact, the results of Table 3 corroborate the hypothesis that the theoretical structure of competencies is coherent with the empirical behavior of the items. In other words, the organization of items by competencies is more relevant to the results of the Enem than the organization of items by subject, which is in accordance with the principles that guide the exam.

It should also be noted that, in general, the correlations in Tables 2 and 3 are relatively small, of low magnitude, when compared to other contexts. However, in the context of ENEM, which is an exam whose score is calculated using the logistic model of the Item Response Theory, it is expected that each item is an independent form of measurement for the same construct, a principle known as local independence (PASQUALI, PRIMI, 2003). In such a context, it is expected that the correlations between the items that measure the same construct will be positive, but low. A very high correlation between two items could indicate that they do not meet the condition of local independence, thus distorting the proficiency estimate, that is, the calculation of the score in each of the four ENEM tests.

## CONCLUSIONS

The internal structure of the set of 175 items was unidimensional in practice, although three sub-factors were identified, in addition to the g factor. The g-factor score showed a correlation of 0.97 with the mean of the 4 tests. The sub-factors showed systematic congruence in 2014 and 2015, with a clear difference between the items of Mathematics and Languages and Codes. In general, the results point to the need to improve the specificity of the constructs related to the tests of Natural Sciences and Humanities.

In the Natural Sciences test, the hierarchical omega coefficient was 0.32 in 2014, revealing serious limitations of the NC score, a single score, to represent the internal multidimensionality of this test – a result compatible with Norris and Phillips' hypothesis about the internal heterogeneity of scientific literacy. On the other hand, the empirical behavior of the items was consistent with the theoretical structure of competencies

described in the ENEM Reference Matrix. In addition, the empirical behavior of the NC items was more coherent with the eight competencies than with the three disciplines, physics, chemistry and biology. In fact, these results confirm the idea that ENEM is structured around general NC competencies, surpassing to some extent the vision of scientific literacy fragmented into disciplines.

# REFERENCES

1.  AERA; APA; NCME. Standards for educational and psychological testing. Washington: American Educational Research Association, 2014.

2.  GOMES, Cristiano Mauro Assis; GOLINO, Hudson Fernandes; DE SOUZA PERES, Alexandre José. Reliability of the scores of the National High School Exam (Enem). Psico, v. 51, n. 2, 2020.

3.  INEP. ENEM reference matrix. Brasília: INEP. 2009.

4.  KOERBER, S.; SODIAN, B.; THOERMER, C.; NETT, U. Scientific Reasoning in Young Children: Preschoolers' Ability to Evaluate Covariation Evidence. Swiss Journal of Psychology, v. 64, n. 3, p. 141–152, set. 2005.

5.  NORRIS, S. P.; PHILLIPS, L. M. How Literacy in Its Fundamental Sense Is Central to Scientific Literacy. Science Education, v. 87, n. 2, p. 224–240, 2003.

6.  PASQUALI, Luiz; PRIMI, Ricardo. Fundamentals of the item response theory: IRT. Psychological Assessment: Interamerican Journal of Psychological Assessment, v. 2, n. 2, p. 99-110, 2003.

7.  PRIMI, R. Intelligence: Advances in Theoretical Models and Measurement Instruments. Psychological Assessment, v. 1, p. 67–77, 2003.

8.  PRIMI, R.; SANTOS, A. a. A. Dos; VENDRAMINI, C. M.; TAXA, F.; MULLER, F. A.; LUKJANENKO, M. D. F.; SAMPAIO, I. S. Cognitive competencies and skills: different definitions of the same constructs. Psychology: Theory and Research, v. 17, n. 2, p. 151–159, 2001.

9.  REVELLE, W. psych: Procedures for Psychological, Psychometric, and Personality Research. Available at: <http://cran.r-project.org/package=psych>. Accessed on: 4 mar. 2024.

10. RODRIGUEZ, A.; REISE, S. P.; HAVILAND, M. G. Evaluating bifactor models: Calculating and interpreting statistical indices. Psychological Methods, vol. 21, no. 2, p. 137–150 (2016).

11. SASSERON, Lúcia Helena; CARVALHO, Anna Maria Pessoa. Scientific literacy: a bibliographic review. Investigations in science teaching, v. 16, n. 1, p. 59-77, 2011.

12. SCHMID, J.; LEIMAN, J. M. The development of hierarchical factor solutions. Psychometrika, v. 22, n. 1, p. 53–61, 1957.

13. TRAVITZKI, R. Evaluation of the quality of Enem 2009 and 2011 with psychometric techniques. Studies in Educational Evaluation, v. 28, n. 67, p. 256, 28 Apr. 2017.