

Applying Content Analysis to Brazilian National Exams

Priscila da Silva Neves Lima¹, Eliene Ribeiro Rosa¹,
Ana Paula Laboissière Ambrósio¹, João Lucas dos Santos Oliveira¹

¹Instituto de Informática – Universidade Federal de Goiás
Goiânia – Goiás – Brasil

priscila_lima@ufg.br, apaula@inf.ufg.br, {ln.ribeiro.rosa, lucas.xms}@gmail.com

Abstract. *Brazil has an educational evaluation system that covers all levels of education. Each level of education has its own system that collects data from the schools, and applies exams and questionnaires to verify the academic level of the students. This paper applies a methodology for the analysis of exams by content knowledge, to two national Brazilian exams: ENEM and ENADE, in order to verify its applicability and adaptation needs. A prototype was implemented to understand the challenges in automatizing the analysis. The methodology showed flexibility and adaptability and shows potential to be adapted to any evaluation exam.*

Resumo. *O Brasil possui um sistema de avaliação educacional que abrange todos os níveis de ensino. Cada nível de educação tem seu próprio sistema, que coleta dados das escolas, e aplica exames e questionários para verificar o nível acadêmico dos alunos. Este artigo aplica uma metodologia para a análise de exames por conhecimento de conteúdo, a dois exames nacionais brasileiros: ENEM e ENADE, permitindo uma comparação entre as duas aplicações da metodologia. Um protótipo foi implementado para entender os desafios em automatizar a análise. A metodologia mostrou flexibilidade e adaptabilidade e mostra potencial para ser adaptada a qualquer exame de avaliação.*

1. Introduction

Educational assessment systems collect and store detailed information, concerning tests, students and their results, and generate reports that inform and sustain decisions at several levels [Dias et al. 2006]. In Brazil, the National Institute of Educational Studies and Research Anísio Teixeira - Inep promotes studies, research and evaluations on the Brazilian Educational System with the objective of subsidizing the formulation and implementation of public policies for the educational sector based on quality parameters and equity, as well as producing clear and reliable information to managers, researchers, educators and the general public. The evaluation system implemented by Inep covers all levels of education. It collects a large amount of data, related to tests, students and performance, data that are not always explored in the official evaluation. The analysis made available by Inep generally refer to descriptive data statistics that describe and summarize the information collected. This information is used to subsidize changes in curricula, teaching methodologies, concepts, training practices, management, among others [Da Silva 2012].

Each level of education has its own system that collects data from the schools regarding structure and staff, including teachers, and applies exams and questionnaires to

verify the academic level of the students. Students finishing high school take the Enem exam, which in addition to verifying students achievements is also used by most public universities as a selection criteria. The Sinaes system (Sistema Nacional de Avaliação da Educação Superior) evaluates higher education. Part of this evaluation is the Enade exam, taken by students beginning and/or finishing undergraduate courses.

Grades are often given for specific parts of the exams, but content analysis of exams, that analyze how students did regarding the general topics covered by the exams, are not always explored in official evaluations, often due to the difficulty of generalizing analysis procedures and standardizing a methodology applicable to all domains. Eventually, punctual content analysis are made by researchers interested in perfecting specific exams [Razmjoo e Tabrizi 2010] [Expert 2018], [Hypolito 2016], in improving pedagogical projects [Costa e Martins 2014], in training teachers [De Lara 2007] [Novossate 2010] [Schwengber 2013], but do not present a generalized methodology and are not part of a systematized analysis.

Lima [Lima et al. 2018] proposes a methodology that allows classification of exam questions into knowledge domains represented in tests and identify analysis that can be performed using this classification. The methodology applies general procedures that allow it to be implemented by a software system. The mechanism separates student's results by themes, that represent general topics covered in the exams, and generates reports that provide a better understanding of the test structure and student outcome by knowledge domain. The information provides decision makers with a new perspective that can have a wide impact, as procedures can be defined for reporting results that allow to understand the student's performance based on the domain knowledge represented in the exam.

In this study we will apply the methodology proposed in [Lima et al. 2018] to the ENEM and ENADE exam, to verify its applicability to national exams and the effort required to adapt the methodology. The article is structured in 6 sections. It starts by presenting the ENEM exam in section 2, followed by a description of the Enade exam in section 3, to allow a comparison between the two applications of the methodology. Then, the methodology for content analysis is discussed. Section 5 shows the applicability of the methodology using ENEM and ENADE data, giving rise to the implemented prototype. Finally, conclusions on the application of the methodology were discussed.

2. ENEM

The National Exam of Upper Secondary Education (ENEM) was created to assess individuals. It consists of four tests, encompassing several fields of knowledge around which basic education pedagogical activities are organized in Brazil, plus an essay [Brasil 2018]. The exam covers general knowledge domains: Human Sciences and their Technologies (History, Geography, Philosophy and Sociology); Natural Sciences and their Technologies (Chemistry, Physics and Biology); Languages, Codes and their Technologies (Portuguese Language, Literature, Foreign Language – English or Spanish, Arts, Physical Education and Information and Communication Technologies); Mathematics and their Technologies (Mathematics). Each test has 45 multiple-choice questions [Silva et al. 2014].

Participants' proficiency is calculated based on answers to the multiple choice test questions using Item Response Theory (IRT). To calculate the average proficiency

by knowledge domain, a simple average of the performance of students participating in the exam by school is used, taking into account those who did the tests and the essay, obtaining proficiency above 0 (zero) in all tests [Brasil 2014]. A report is given to each student containing five independent grades, one for each major area covered by the exam and one for the essay test. Grades go from 0 to 1000.

3. ENADE

The National Assessment of Student Achievement (ENADE) – is an exam used in the assessment of undergraduate programs in higher education institutions throughout Brazil. It consists of an exam administered to students who are entering or finishing their undergraduate courses. The programs are grouped in three representative areas (Agriculture/Health, STEM, Social Sciences) and each year one group is assessed, meaning that programs in these areas are assessed every three years. Each program has its own test, with domain specific questions to evaluate how students perform nation wide, which allows comparisons between equivalent courses offered in different institutions.

ENADE is comprised of two instruments: a socioeconomic questionnaire, used to define student profile, and a test, composed of multiple-choice and essay type questions, divided into two parts. The first part, "general formation (GF)", presents itself as a "common component" to the tests of the different domains being evaluated in that year, investigating general skills and knowledge expected from undergraduate students independent of their course. The second part, "specific component (SC)", contemplates the specificity of each domain, both in the field of knowledge and in the skills expected for the professional profile. The specific component of the tests is defined by the guidelines and matrices prepared by the domain's assessment advisory committee and the ENADE general assessment advisory committee. These matrices comprise content in accordance to the national curricular guidelines for undergraduate courses.

The test structure is the same for all courses and contains 40 questions: 10 general formation questions (2 being essay type questions) and 30 specific component questions (3 being essay type questions). Questions with low discrimination power are annulled and are not used in the grading process [Brasil 2005]. In addition, the calculation of the grades also considers the nature of the question, assigning different weights to multiple choice and essay type questions. Using these percentages, ENADE generates, for each student, one grade for GF and one grade for SC. Student's final score is the weighted average obtained in each part, with 25% for the general formation, and 75% for the specific component [Santos et al. 2017].

4. Content Analysis

To perform exam content analysis one must understand the exam structure and the domain knowledge being verified by the exam. Often, widely applied tests release documents stating the content that will be evaluated, as well as its structure and type of questions. Information is also given as to how scores are calculated, including maximum score for each question and weight in the final grade. The methodology proposed by [Lima et al. 2018] takes into account all these factors and can be applied to any large-scale assessment. It is structured in three phases detailed below.

Phase 1 - Identification/Cataloging

This phase includes the identification of the exam questions to be analyzed and their classification in previously defined themes. These themes should be non-overlapping and complementary subsets of the programmatic content of the exams, or may be created by domain knowledge experts based on their analysis interests. These themes correspond to the sub-domains in which analysis will be undertaken. Therefore, they must be specific enough to be representative of the important content domains, but generic enough to allow several questions do fall into that theme.

Phase 2 - Group by theme

In this phase, a grade for each theme is calculated for each student. This is done by identifying the student's result for each exam question, grouping the questions by theme and calculating a score obtained by that student in that theme. If questions have different weights in the test's final grade, it may be advisable to maintain these weights when calculating the scores by theme. How this grade is calculated can be defined according to the situation, but they must be normalized as to allow comparison between themes. Other aspects of data transformation must also be taken into account, such as missing values and outliers. If the dataset has not been pre-processed to treat these aspects, this must be done before proceeding to calculating the score by theme.

Phase 3 - Analysis and results

In this last phase, the data is already transformed and grouped. Data analysis techniques such as descriptive and inferential statistics can be applied [Howell 2010]. Another powerful analysis technique, data mining proposes to abstract relevant information from large volumes of data, allowing to establish relationships and interpretations that serve as the basis for the construction of new knowledge [Kulkarni et al. 2013]. Data mining techniques also provide predictability of the outcome of a future observation. What type of analysis is best will depend on users' expectations and the available data.

5. Applying the Methodology

The methodology was applied according to the general structure of section 4, taking into account characteristics of the national exams. Each phase was subdivided into three activities each (Figure 1). All data used in this research are open and made available by Inep¹.

¹Available in: <http://www.inep.gov.br/>. Accessed in: 03/25/2018.

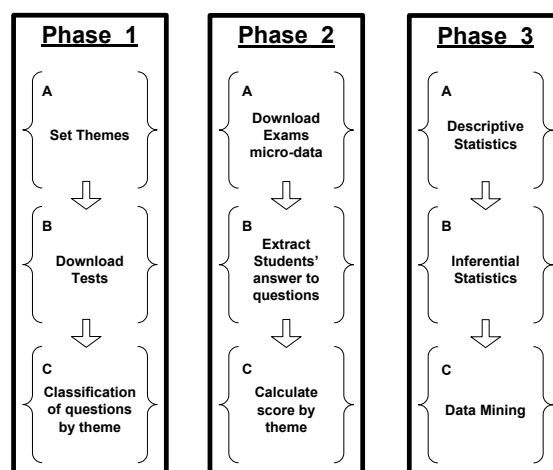


Figura 1. Methodology phases applied to Inep exams

To help study the application of the methodology to the Inep exams, two study cases were selected. For ENEM, data from the 2017 exam was analyzed. For ENADE, data from the specific component of the 2014 Computer Science exam was analyzed.

A proof of concept prototype was developed to verify the viability of automating the analysis process. It was implemented in Java, using JFreeChart3, Commons Math4 and JRI5 to do the statistical analysis.

Phase 1 - Identification/Cataloging

ENEM evaluates high school student's domain knowledge (e.g. Portuguese, Mathematics, Geography, History, etc). The set of questions for each year is the same for all students, however, there are six versions of the test where the order of the questions are different. Thus attention must be paid to the version taken by the student in the classification step.

ENADE assesses undergraduate students from different courses. As discussed before, the test is divided between general knowledge and domain knowledge, since ENADE seeks to evaluate mainly the academic area of the student (e.g. Computer Science, Psychology, Engineering, Medicine etc). There is a set of common questions to all exams in a given year (general formation), and a set of questions exclusive to the course being evaluated (specific component). Therefore, themes can be defined for the common component which can be analyzed for all exams in a given year, or for the specific component, in which case themes will be defined for each course exam.

This difference in the structure of the exams brings a diversity in the analysis of ENADE. If the researcher wants to analyze the general knowledge, it is possible to consider all years of application of the test. Otherwise, if the researcher wants to analyze a specific course it is necessary to consider the years when the application of the test was performed for this particular course.

However, the most important is that although there are differences between the exams like the structure of the tests, the volume of information, the level of the students evaluated etc., the steps taken in the applied ENEM methodology are the same to the methodology applied to ENADE.

Step A) Set Themes

Eleven themes were identified and used in the cataloging process of the ENADE exam in Computer Science: (1) Algorithms, Data Structure and Programming; (2) Computer Architecture and Digital Circuits; (3) Database; (4) Computer Graphics and Image Processing; (5) Software Engineering; (6) Ethics, Computer and Society; (7) Artificial Intelligence; (8) Formal Languages and Automate, Compilers and Computability; (9) Mathematical Logic, Discrete Mathematics, Statistics and Graphs; (10) Computer Networks, Distributed Systems and Telecommunications; and (11) Operating Systems.

For ENEM, 14 themes were identified and used in the classification process: (1) History; (2) Geography; (3) Philosophy; (4) Sociology; (5) Chemistry; (6) Physics; (7) Biology; (8) Portuguese Language; (9) Literature; (10) Foreign Language – English or Spanish; (11) Arts; (12) Physical Education; (13) Information and Communication Technologies; (14) Mathematics. In addition, students write an essay test.

Both ENADE and ENEM themes were defined based on the content matrices for the courses that have remained constant for several years. This means that the themes, once defined, can be used for subsequent tests with little or no change.

The definition process was quite straight forward as the content for these tests are clearly defined and organized by levels of abstraction. As to obtain themes that would aggregate several questions, higher levels of abstraction were used. However, the list of themes is intimately related to the analysis that will be undertaken. Therefore, this list can be defined using other criteria, as long as the themes are clearly defined. Questions will be allocated to only one theme.

Step B) Download Tests

ENADE exams are available in the Inep/ENADE website ², and were downloaded from the Inep/ENEM website ³ for ENEM. Both exams have tests in Portable Document Format (PDF) format.

As discussed before, the ENEM has 6 (six) different versions of the same exam. However, Inep standardizes the results adopting the question order of version 1 of the exam, ie, even-though the students answered the questions in different orders, the answer vector in the micro-data follows the order of version 1 of the exam. Therefore only version 1 of the tests need to be downloaded and classified, remembering that parts of the ENEM exam are taken on different days and there are more than one test involved.

Step C) Classification of questions by theme

In both cases, each question was associated to a theme. Association was performed by reading the question and linking it to a theme. The result of this classification process is

²Available in: <http://portal.inep.gov.br/web/guest/provas-e-gabaritos3>. Accessed in: 03/25/2018.

³Available in: <http://portal.inep.gov.br/web/guest/provas-e-gabaritos>. Accessed in: 11/03/2018.

a file containing for each question, the year of the exam, the question, its number in the exam, the theme to which it was associated, and if it was used in the final score or if it was annulled. If necessary, additional information may be included in the file.

Phase 2 - Group by theme

This phase of the methodology, despite having the same applicability for both tests, was more costly for ENEM due to the amount of data. In 2017, the ENEM edition had approximately 6.7 million subscribers [MEC 2018]. Manipulating this data computationally and generating the correct response vector required high computational performance. However, generating the score by theme was more difficult for ENADE. Because of the essay questions and the weights associated to the questions.

Step A) Download micro-data

Inep micro-data constitute the lowest level of data desegregation collected by surveys, evaluations and examinations. The information is available in ASCII format, and contains inputs for reading through statistical softwares. These files contain data for all students who completed ENADE and ENEM in that year. ENADE and ENEM micro-data was downloaded through the Inep⁴ website. The micro-data provided by Inep is standardized which facilitates manipulation. This standardization allows the use of this methodology and also allows it to be applied to other Inep exams, enabling the perspective of a unique automatic system of educational analysis.

The files contain data on all students that registered for the tests. However, many did not take the tests or did not complete them. For the analysis, only completed tests were taken into account. To identify completed tests, one of the variables contained in the micro-data was used. For both exams, students that did not answer the social-economic questionnaire were also excluded. This way, entries with missing values were eliminated.

Furthermore, in the case of ENADE, the downloaded worksheet contains data on all students that took the test on that given year. Therefore, to analyze data from the Computer Science course, it was necessary to use the information contained in the worksheets to identify the domain and sub-domains, where appropriate, of the student's course, selecting the data relevant to our analysis. For ENADE, the number of participants in the 2014 Computer Science course that completed the exam and the questionnaire was 8266 students.

Step B) Extract students' answer to questions

The ENADE micro-data worksheets contain a column with vectors that indicate the correctness of the multiple choice part of the specific component questions in the exam. Each position in the vector identifies a specific component question in the exam. The number of the question in the exam identifies its position in the vector. The possible values for each position in the vector are: 0 = Wrong, 1 = Right, 8 = annulled by the commission, 9 = annulled by the discrimination index. When a student does not answer a certain question,

⁴Available in: <http://portal.inep.gov.br/microdados>. Accessed in: 12/24/2018.

it is considered a wrong answer, that is, 0 is assigned to the question. Essay questions have separate columns containing the score for that question. Scores vary from 0 to 100. To extract the students' response to questions it was necessary to dismember this vector, group the questions that belong to the same theme and calculate the grade by theme for each student.

The ENEM micro-data worksheets do not contain a column with vectors that indicate the correctness of the multiple-choice questions in the exam. These worksheets contain the students' answers and a correct answer vector. It was necessary to create a vector like the one in ENADE that indicates if the student got a question right or not (hit) and then dismember it. In order to create the hit response vector, the vector of responses was confronted with the correct answer vector. When they matched, it was assigned 1 and otherwise, assigned 0. To annulled questions a 9 was assigned. When a student does not answer a certain question, it is considered a wrong answer, that is, 0 is assigned to the question. The Essay test has a separate column containing its score. Scores vary from 0 to 1000.

Step C) Calculate score by theme

Having identified the student's score for each multiple-choice question in the exams, it is possible to calculate a raw score for each theme in these exams, adding the number of questions associated with the theme that they answered correctly and dividing by the number of questions in that theme. This information is stored in a new column added to the worksheet. Annulled questions are not considered in this calculation.

In ENADE, another column was used to store the student's average for the essay questions. If no essay question was associated to a theme, the value 0 was attributed to all students. Finally a column was also added containing the final score for that theme, calculated using the weighted average considering 85% for the objective/multiple-choice questions and 15% for the essay questions, adopting the same proportion used to calculate the final score in the exam. As ENEM only has objective questions, to generate the score by theme it was enough to group the questions by theme and calculate the grade. The essay test is not classified by theme and the grade is assigned in a separate column.

This generates a general worksheet where each line contains the year of the exam, selected information about the student from the original micro-data file, and information about theme scores. This file compiles the information obtained from all the downloaded exams, including socio-economic data, institution where the student did the course, etc.

Phase 3 - Analysis and results

For ENADE, the report⁵ provided by Inep for the 2014 exam was also downloaded. These reports contain statistical analysis of students' profile and outcomes. Based on these results, facility and discrimination indexes are calculated for each question and detailed in the reports. These reports provided insight into what specialists expected in a report and how information could be described to complement this report. ENEM however does not provide updated reports. The only report available was from 2011-2012.

⁵Available in: <http://portal.inep.gov.br/web/guest/relatorios>. Accessed in: 12/24/2018.

There are several techniques for analyzing educational data that can be applied, such as *Descriptive Statistics*, that will be used to analyze the exam structure and student outcome; and *Inferential Statistics* that allows analysis of correlation between variables, differences between groups or differences in different moments in time [Martins 2011]. *Data Mining* will be used mainly to predict outcomes.

This phase generates a report containing the results obtained from the different analysis procedures applied to the transformed data. It is hoped that the analysis and results presented complement other reports provided by Inep and together subsidize political-pedagogical redefinitions to the training paths in the Brazilian educational scenario.

6. Conclusion

This paper applies the methodology for the analysis of exams by content knowledge, proposed by [Lima et al. 2018], to two national Brazilian exams: ENEM and ENADE, to verify the complexity of the approach. A prototype was implemented to understand the challenges in automatizing the analysis. The methodology showed flexibility and adaptability and shows potential to be adapted to any evaluation exam.

The implementation showed that phases 2 and 3 of the methodology can be almost completely automated. The only part that still required human intervention was the download of the micro-data. However, once this is done for all the required data or direct access is given by Inep, these phases will be automated. Phase 1 was the more human demanding. Although the definition of the themes was done based on the content matrices used in the exams, the list was defined based on human expertise. This is also true for the classification of the exam questions within a theme. Implementing an automatic approach to phase 1 would require other technologies including text analysis.

Reuse of software to implement the analysis of both tests was easy due to the use of similar rules and patterns used by Inep in the definition and creation of the tests and resulting data. Furthermore, Inep micro-data has already been pre-processed, one of the more demanding activities in the data analysis process, requiring only filtering of the desired data, and manipulation of the result vectors to calculate grades for each theme.

Although this is an exploratory study of the methodology, results encourage a more extensive application not only of the implementation aspects, but also of the results obtained from the exam content analysis. It would be interesting to apply the methodology to all exams offered by Inep as some themes, such as portuguese and mathematics, appear in several exams which would allow longitudinal studies. For this, strategies would have to be defined to join these results as the data is anonymous and it is not possible to identify records of exams taken by the same person.

Referências

- Brasil (2005). Relatório de Síntese 2005. *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira*.
- Brasil (2014). Nota Explicativa Enem 2014 Por Escola. *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira*.

- Brasil (2018). Basic Education Assessments. *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira*. Available in: <http://portal.inep.gov.br/web/guest/basic-education-assessments>. Accessed in: 20 nov. 2018.
- Costa, J. P. D. C. e Martins, M. I. (2014). O ENADE para a licenciatura em física: Uma proposta de Matriz de Referência. *Revista Brasileira de Ensino de Física*, 36, n. 3(3):3401 – 3401/9.
- Da Silva, J. C. (2012). ENADE: Um Estudo Sobre o Desempenho das Instituições de Ensino Superior nos Anos de 2007 e 2010. *Dissertação (Mestrado) - Universidade Católica de Brasília*.
- De Lara, I. C. M. (2007). Exames Nacionais e as "verdades" sobre a Produção do Professor de Matemática. *Tese (Doutorado) - Universidade Federal do Rio Grande do Sul*.
- Dias, C. L., De Lourdes, M., Horiguela, M., Marchelli, P. S., São, U., e Resumo, M. (2006). Políticas para avaliação da qualidade do Ensino Superior no Brasil: um balanço crítico. 32(3):435–464.
- Expert, M. (2018). XAT 2018 Analysis: In-depth XAT Exam Analysis.
- Howell, D. C. (2010). *Statistical Methods for Psychology*. Belmonte: Wadsworth, 7^a ed. edition.
- Hypolito, V. A. H. A. (2016). Uma análise do conteúdo das provaas da OBMEP Nível 3. *Technical report*.
- Kulkarni, M. G., Rampure, M. C., e Yadav, M. (2013). Understanding Educational Data Mining. *International Journal of Electronics and Computer Science Engineering*, 2, n.2:773–77.
- Lima, P. d. S. N., Ambrósio, A. P. L., Félix, I. M., Brancher, J. D., e Ferreira, D. J. (2018). Content Analysis of Student Assessment Exams. *48th Annual Frontiers in Education (FIE)*.
- Martins, C. (2011). *Manual de análise de dados quantitativos com recurso ao IBM SPSS*.
- MEC (2018). Relatório de Gestão do Exercício de 2017 do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - Inep. Technical report.
- Novossate, S. (2010). O ENADE e os Documentos Curriculares: Um Estudo sobre a Formação de Professores de Biologia. *Dissertação (Mestrado) - UFPR*.
- Razmjoo, S. A. e Tabrizi, H. H. (2010). A Content Analysis of the TEFL M . A . Entrance Examinations (Case Study : Majors Courses) Seyyed Ayatollah Razmjoo and Hossein Heydari Tabrizi. *Applied Linguistics*, 14(1):159–170.
- Santos, R. A., Quintilio, R., Vargas, S. M., e Teixeira, U. (2017). Nota Técnica N 2/2017/CGCQES/DAES. *INEP*, páginas 1–7.
- Schwengber, L. (2013). Exame Nacional de Desempenho de Estudantes: Problematisando Verdades sobre a Formação do Professor de Matemática. *Dissertação (Mestrado) - Universidade de Santa Cruz do Sul*.
- Silva, L. A., Morino, A. H., e Sato, T. M. C. (2014). Prática de Mineração de Dados no Exame Nacional do Ensino Médio. *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, 3(1):651–660.