iscte

INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Neural Network Approach for Question Generation Using the Revised Bloom's Taxonomy

Gonçalo Fernando Ferreira da Costa Durão Correia

Master in Computer Engineering

Supervisors:
Doctor Ricardo Daniel Santos Faro Marques Ribeiro, Assistant Professor,
ISCTE-IUL

Master Hugo Patinho Rodrigues, Researcher
INESC-ID Lisboa

November, 2020

iscte

TECNOLOGIAS
E ARQUITETURA

Neural Network Approach for Question Generation Using the Revised Bloom's Taxonomy

Gonçalo Fernando Ferreira da Costa Durão Correia

Master in Computer Engineering

Supervisors:
Doctor Ricardo Daniel Santos Faro Marques Ribeiro, Assistant Professor,
ISCTE-IUL

Master Hugo Patinho Rodrigues, Researcher
INESC-ID Lisboa

November, 2020

Acknowledgments

Throughout the elaboration of this dissertation and through all the obstacles and challenges it brought me as a student, there are several people that were vital in making this document a reality. As such, i would like to begin by thanking them.

First and foremost, i would like to thank my Supervisor, Dr. Ricardo Ribeiro, for the assistance, comprehension and guidance throughout the constitution of this document. I would also like to thank my Co-Supervisor, Hugo Rodrigues, for the endless support, constructive criticism and objectivity that is nothing short of impeccable and key to the elaboration of this document.

On a more personal note, I wish to thank my family for granting the necessary stability and wisdom that allowed me to shift my focus to this document even throughout the most complicated moments.

Finally, a thank you to all my friends who actively motivated me to fulfill this objective.

Resumo

Questionar é uma parte fundamental do processo de aprendizagem. À medida que novos conteúdos surgem e se torna vital a sua compreensão para a sociedade moderna, a geração de questões torna-se uma necessidade que, quando feita manualmente, requer tempo e recursos para ser eficaz. Neste documento introduzimos uma abordagem Sequence-To-Sequence (Seq2Seq) que consiste na geração de uma variedade de questões relevantes para os contextos nas quais são colocadas. De forma a garantir que as questões geradas são diversas, relevantes e de valor acrescentado para situações de aprendizagem, utilizámos a Taxonomia de Bloom Revista (TBR), uma taxomia de aprendizagem que é orientada aos objetivos da aprendizagem e pode ser utilizada para separar questões com base no seu nível cognitivo. Contudo, os modelos de redes neuronais precisam de grandes conjuntos de dados para o seu treino e os datasets atuais orientados à TBR são pequenos e escassos. Para colmatar esta falha, desenhámos um classificador de questões a ser usado para categorizar atuais e futuros datasets tendo em conta as orientações da taxonomia. Utilizámos este classificador para criar um dataset posteriormente utilizado para treinar o modelo Seq2Seq proposto. Adicionalmente, para cobrir os diferentes níveis da taxonomia, criámos seis modelos *fine-tuned* específicamente para cada um dos níveis cognitivos da TBR. Os resultados mostram que a nossa abordagem é promissora, garantindo variedade de questões para todos os níveis da taxonomia, ultrapassado a baseline quando avaliada usando BLEU-1, e considerada por avaliadores humanos, de forma geral, como uma abordagem que produz questões bem escritas, relevantes e compreensíveis.

Palavras-chave: Geração de Questões, Taxonomia de Bloom Revista, Classificação de Questões

Abstract

Questioning is a fundamental part of the learning process. As new content arises and learning it becomes vital to the modern society, *question generation* becomes a necessary job that requires time and resources to be performed effectively. In this document, we propose a Seq2Seq approach that generates a variety of questions that are relevant to the contexts where they are asked. In order to ensure that the generated questions are diverse, relevant, and valuable to learning situations and environments, we use the Revised Bloom's Taxonomy (RBT), a learning taxonomy that is oriented to learning objectives and can be used to separate questions based on their required cognitive level. However, neural network models require large collections of data to be trained, and datasets addressing RBT are small and scarce. To address this gap, we designed a question classifier that can be used to label current and future datasets using the guidelines provided by RBT. We employed this classifier to create a labeled dataset, which was then used as training data for our proposed Seq2Seq model. In addition, to cover the different taxonomy levels, we create six different fine-tuned models aimed specifically to each one of RBT cognitive levels. Results show that our approach is promising, guaranteeing a variety of questions for all levels of the taxonomy, surpassing the baseline when measured by BLEU-1, and deemed overall well-written, relevant and understandable, by human evaluators.

Keywords: Question Generation, Revised Bloom's Taxonomy, Question Classification

Contents

List of Figures

List of Tables

Acronyms

AB AdaBoost.

Adam Adaption Moment Estimation.

BERT Bidirectional Encoder Representations from Transformers.

BiLSTM Bidirectional Long Short-Term Memory.

BOW Bag-of-Words.

BT Bloom’s Taxonomy.

CGC-QG Clue Guided Copy Network for Question Generation.

CNB Complement Naive Bayes.

GB GradientBoosting.

GCN Graph Convolutional Networks.

GloVe Global Vectors.

GRU Gated Recurrent Unit.

k-NN K-Nearest Neighbours.

KB Knowledge Base.

lbfgs Limited-Broyden–Fletcher–Goldfarb–Shanno.

LR Logistic Regression.

LSTM Long Short-Term Memory.

MCQ Multiple Choice Question.

MNB Multinomial Naive Bayes.

MOOC Massive Open Online Course.

NB Naive Bayes.

NER Named Entity Recognition.

NLP Natural Language Processing.

NN Neural Network.

OS OpenStax.

POS Part of Speech.

QA Question Answering.

QC Question Classification.

QG Question Generation.

QuAC Question Answering in Context.

RBT Revised Bloom's Taxonomy.

ReLU Rectifier Linear Unit.

RMSprop RMSprop.

RNN Recurrent Neural Network.

SDT Syntatic Dependency Tree.

Seq2Seq Sequence-To-Sequence.

SGD Stochastic Gradient Descent.

Sklearn Scikit-Learn.

SMOTE Synthetic Minority Oversampling Technique.

SQuAD Stanford Question Answering Dataset.

SQuAD 2.0 Stanford Question Answering Dataset 2.0.

SRL Semantic Role Labeling.

SV Subject-Verb.

SVM Support Vector Machine.

SVO Subject-Verb-Object.

1 Introduction

1.1 Motivation

Questioning is a fundamental part of the learning process. By answering questions it is expected that an individual's learning outcome improves [1]. As more information becomes available across different sources of knowledge, manually creating questions that can keep up with all these novelties proves to be a difficult task, requiring time, effort and expertise in order to be done successfully.

1.1.1 Providing an Automatic Solution

When considering the impact that educational assessments have on a student's cognitive development [2], we understand that it is paramount to create questions that adequately challenge the student at different degrees of difficulty. Not everyone learns the same way when exposed to the same questions and most assessments are created generically for a group of students rather than being custom-designed, especially due to the fact that customization would largely increase the expended time in the task of question generation.

When performed automatically, the process of question generation in any learning circumstance allows educators and tutors to focus their time in other complex and necessary tasks by relieving them of their function as a question creator. It would also promote self-learning by providing a tool that would, by receiving information as input, generate quality questions with added value to a learner's experience.

1.1.2 Providing a Bloom's Taxonomy-based Solution

Learning taxonomies function as guidelines for learning objectives. When applied to assessments, generating questions becomes a more accurate mean to measure a student's cognitive skill, which in turn can prompt teachers to tailor their assessments according to specific needs [3].

In 1956, Benjamin Bloom and his team developed the Bloom's Taxonomy (BT) [4], a *learning taxonomy* that classifies the cognitive process under one of six hierarchically structured levels: *remembering*, *understanding*, *applying*, *analysing*, *evaluating* and *creating*, with *remembering* being the lowest level of cognition and *creating* the highest. This taxonomy has later been revised and improved by David Krathwohl [5], giving birth to the Revised Bloom's Taxonomy (RBT).

By manually following a *learning taxonomy*, it is mandatory that the entity who creates the questions knows how to label them beforehand, which would drastically increase the time and effort spent in writing questions while also being error-prone due to personal interpretation, especially in ambiguous cases [6].

1.1.3 Harnessing the Potential of Natural Language Processing

Natural Language Processing (NLP) is the field of study that focuses on the understanding of how the interactions between humans and computers can be represented and processed. Automatizing the process of creating questions while respecting the RBT vastly depends on a set of NLP tasks.

Primarily focused in the creation of questions, Question Generation (QG) is a NLP task that has several applications such as the creation of assistant chatbots that interact with users to obtain domain-specific information [7], the development of educational tools that help students attending Massive Open Online Courses (MOOCs) [8], or even studying individually (self-assessment) [8].

Another important NLP task is Question Classification (QC), which provides means to attribute a class to a given question and can provide support to train question generators or be used by itself (for example in the identification of a question according to school subjects or as a support tool for question answering systems [9]). In the context of this dissertation, QC is quintessential for the labeling of a dataset that can aid the learning process of QG.

1.1.4 Research Gap

To the best of our knowledge, no prior studies of Sequence-To-Sequence (Seq2Seq) *question generation* that rely on the guidelines provided by the RBT were found. We have also not found any publicly available, domain-agnostic dataset that contain questions labeled using RBT (for the purposes of generation or classfication).

1.2 Objectives

The fundamental objective of this dissertation is the proposition of a domain-agnostic *question generation* approach based on *Neural Networks (NNs)* that follows the guidelines proposed by the RBT. Given a text as input as well as the selected level of the taxonomy (or even just the text as input), the system's output is expected to be comprised of questions that abide by the RBT and are therefore suitable for a large variety of learning contexts, especially in assessments and self-learning.

A secondary objective of this dissertation is a direct consequence of the necessity to fulfill the first objective. The current datasets available are not optimal to train a *question generator* and manually labeling a well-established dataset is very expensive time and resource-wise. We propose the creation of a *question classifier* that can be trained with significantly less data and have the resulting model (a RBT question classifier) label an existing dataset for the purpose of training a QG approach. Nevertheless, it is also intended that the question classifier can be used as a standalone component for the classification of user questions according to the cognitive level in the RBT.

1.3 Document Structure

Chapter 2 presents a comprehensive literature review that elaborates on all the previous work pertaining to the development of this dissertation.

Chapter 3 dwells on the classifier portion of the proposed solution, detailing all the choices and results related to this task.

Chapter 4 presents the generator of the proposed solution. It provides insight on all the relevant information regarding the generation of questions, how well it performs, and how it can be adjusted for a specific level of the taxonomy as well as the relevant results pertaining to this task.

Chapter 5 summarizes the dissertation's conclusions and completes the document by presenting the future work.

2 Literature Review

As mentioned in Chapter 1, the goal of this dissertation is to propose an approach for learning environments that takes advantage of the capacities of *Neural Networks (NNs)* and Natural Language Processing (NLP) in order to generate questions that put into practice the guidelines defined in the Bloom's Taxonomy (BT).

In this chapter, Section 2.1 expands on the concepts of BT [4] and Revised Bloom's Taxonomy (RBT) [5], comparing and contrasting both versions and explaining the rationale, as well as the challenges, behind the selected taxonomy for this dissertation. Section 2.2 summarizes several approaches to Question Generation (QG) and presents justification for the usage of *NN* as opposed to other methods of QG. Section 2.3 gives context on commonly used datasets regarding QG and Question Classification (QC) while also providing considerations on the current dataset scenario, stating the dataset choices in the scope of this dissertation. Finally, Section 2.4 addresses previous work in the task of QC and reaffirms the underlying objective that such task has in this dissertation.

2.1 Bloom's Taxonomy

Quality question writing is a process that requires time and effort from tutors, lecturers and examiners. Considering the definition of learning objectives (referenced by Näsström [10] as standards) as "what a student should know or should be able to accomplish" [10], *learning taxonomies* are considered to be a useful tool not only for the classification of learning objectives in well-defined categories, but also to study how learning objectives change over time [10].

BT was originally intended as a tool to reduce the necessary effort in the production of annual comprehensive examinations [5]. It has also been concluded that it can play a large role in the classification of both learning and teaching objectives in educational environments [11, 12, 10].

This taxonomy is built upon three learning domains: *cognitive*, *affective* and *psychomotor*, each with a specific purpose. The *cognitive* domain considers what you can learn in the form of intellectual skills. The *affective* domain considers the learning and development of subjective concepts such as feelings and values. Finally, the *psychomotor* domain attends to the development of physical capabilities. From all three domains, the *cognitive* is the only one pertaining to the development of intellectual skills and knowledge, therefore being the only domain in the scope of this dissertation.

Throughout this section, the *cognitive* domain of BT will be thoroughly discussed based on prior work, as well as the changes introduced by RBT. Finally, the challenges posed by this taxonomy in the proposed approach will be presented.

2.1.1 Bloom's Original Taxonomy

Bloom's Original Taxonomy of Educational Objectives [4] dates back to 1956 and its *cognitive domain* is comprised of *six* levels that are hierarchically organized in ascending order of complexity:

- *Knowledge* – The simplest level in terms of complexity, placed at the bottom of the hierarchy. This level of complexity emphasizes the ability to remember [4]. It is achievable by simply memorizing facts and recalling previously obtained knowledge. Asking to list facts or to define concepts that were previously taught are two examples of questions that require this level of complexity.

- *Comprehension* – This level goes a step above by imposing the necessity of understanding how a given concept works. Bloom et al. [4] define behaviours of this cognitive level to be interpretation, translation and extrapolation. Asking to explain a concept rather than simply writing a formal and previously memorized definition is a way to test for this level of complexity.

- *Application* – At this level, context specific problems are introduced by providing example scenarios which can only be answered by applying prior knowledge to the said scenarios [12, 13]. Bloom et al. [4] further differentiate the concepts of comprehension and application by explaining the former as using an abstraction when prompted to do so, while the latter refers to correctly applying an abstraction in a situation without any cues on how to do it in the given problem.

- *Analysis* – Bloom et al. [4] state that this level requires the ability to divide information into simpler parts and detect their relationships. Analytical level questions may include processes such as drawing relationships between concepts, creating assumptions or classifications based on the characteristics of the concepts [13]. Asking how a process occurred and how is it influenced by given circumstances is an example of a possible question at this level.

Table 2.1: Bloom's Taxonomy Cognitive Process Dimension Overview

Cognitive Level	Original Bloom's Taxonomy	Revised Bloom's Taxonomy
1	Knowledge	Remembering
2	Comprehension	Understanding
3	Application	Apply
4	Analysis	Analyze
5	Synthesis	Evaluate
6	Evaluation	Create

- *Synthesis* – The second highest level in terms of complexity requires the ability to build something new (a plan, product or proposal) by combining previous learned components [13, 14]. Bloom et al. [4] refer to *synthesis* as the putting together "elements and parts so as to form a whole". Asking to build a new product based of an already known set of materials facilitates the assessment of this level.

- *Evaluation* – To evaluate information is the most complex activity that can be observed in BT. It relates to the ability to make judgements, criticize, appraise, and defend a point of view about a given subject or piece of information [4, 13].

2.1.2 Revised Bloom's Taxonomy

Unlike the original BT, where the emphasis of the cognitive hierarchy relied on nouns, RBT focuses on actions which are expressed by verbs. Originally, the knowledge level of BT would unidimensionally allow for both verb and noun interpretations which is considered by Krathwohl [5] as an anomaly that had to be fixed. RBT considers objectives as a set of *two dimensions*: the content (noun) and what is supposed to be done with such content (verb). In RBT, the noun makes way for the *Knowledge dimension* while the verb originates the *Cognitive Process dimension*.

When compared to the *Cognitive Process dimension* of the original BT, three levels were renamed, two switched places in the hierarchy and all of the levels are now written as verbs. Despite these alterations, the main definition of each level retains its essence and can be consulted in Section 2.2.1. Table 2.1 facilitates the comprehension of the differences between BT and RBT regarding the *Cognitive Process dimension*.

Table 2.2: Revised Bloom's Taxonomy Knowledge Dimension Categories.

Knowledge Category	Definition
Factual Knowledge	Refers to what basic elements are needed in order to know how to solve problems and work with a given subject, such as terminology.
Conceptual Knowledge	It is defined by the relationships between basic elements and what can derive from them, such as theorems, principles and classifications.
Procedural Knowledge	Knowledge pertaining to the domain of methods and techniques, usage of skills including when to use them.
Metacognitive Knowledge	The most abstract of the presented knowledge, refers to cognition itself and the self-awareness of that cognition.

The *Knowledge dimension* of the original taxonomy was defined in three categories but without a clearly defined separation. The *revised knowledge dimension* allows for new ways to evaluate objectives by having a fourth category created – *Metacognitive Knowledge*, which is related to an individual's self-awareness of the fact that they are thinking or learning. An overview of the revised taxonomy's knowledge categories is presented in Table 2.2.

The duality of dimensions in the revised taxonomy results in an increased complexity in evaluation processes. Because it has now two dimensions, it is represented using a table where the *Knowledge dimension* can be perceived in the rows and the *Cognitive dimension* in the table's columns. Lee et al. [7] state that this table is a viable system to classify course objectives with clarity and provides insight on that same course.

Research was also conducted in the scope of RBT and its usage in different situations: Näsström [10] studies the usefulness of this taxonomy when applied to mathematical learning objectives (by utilizing the taxonomy table as a tool, while also providing a comparison between teachers and assessment experts when interpreting those standards. Concerning the role of the taxonomy in this research, the table was deemed as useful (at least for the defined mathematical earning objectives) due to the fact that it had almost all cells filled by all the judges in the study.

Shah et al. [8] explore the impact of the taxonomy in assessment by examining units and organizing their objectives, assessments and instructional activities taking into consideration the guidelines provided by the table. This research concludes that by focusing on the taxonomy table, assessment can be aligned with instruction and objectives. Another relevant conclusion of this research is that due to the nature of the taxonomy table being learning-centered instead of performance-centered, the main concerns regard the *cognitive processes* and *knowledge processes* that allow students to reach learning objectives and not the specific knowledge of items included in such assessments.

2.1.3 Challenges Using The Original and Revised Taxonomies

To the best of our knowledge, *question generation* alongside RBT was not studied yet. This may be due to the fact that RBT presents emphasis in assessment and definition of learning objectives, leaving the teaching component to classroom teachers. Also, the complexity added by the knowledge domain could add undesirable difficulty in creating classroom examinations as the documents that could explore the table in full would require a minimum of 24 questions (not considering questions that may target more than one table cell).

When evaluating a question according to this taxonomy, there is a certain degree of subjectivity and opinions may differ according to the individual that classifies it. Most of these subjective episodes occur in questions that have an action verb that can be placed in more than one level of the taxonomy. A simple solution would be to classify this question with the highest level it can obtain. This methodology would be possible because there is a hierarchically dependency relationship between categories, for example, in order to *apply* a concept, a student must *remember* and *understand* that concept.

BT and RBT are very similar when considering exclusively their *cognitive domain*, with the only practical differences being the exchange in places between *evaluation* and *synthesis* and the translation of nouns into more representative verbs. The hierarchy presented in RBT is the selected hierarchy to be used throughout this dissertation, as the revision pertains to a more recent date and it is considered in the scientific community as an improved version of the cognitive domain.

The *Knowledge dimension* will be left out from this dissertation due to the increased challenge regarding classification expertise.

Finally, a major obstacle to the creation of a viable solution is the lack of a vast, agreed upon dataset based on BT or RBT. This matter will be thoroughly discussed in Section 2.3 where issues directly related to datasets are exposed.

2.2 Question Generation Overview

Generating questions as an automatic task has been researched using various different strategies and tools. Although many approaches exist, Shah et al. [8] define them as being mostly rule-based and template-based. Hybrid solutions combining the previous two along with other methods have also been thoroughly studied. With the development of NN, researchers have also explored the usability of these mechanisms in QG. In this section, all the aforementioned approaches will be presented as well as a review of related work.

2.2.1 Rule-based Approaches

Rule-based processes can generically be characterized by having the necessity of designing ways to deal with the *inter-dependency* of words in a sentence (essentially how words are related and how that relationship profiles specific kinds of sentence).

Heilman and Smith [15] utilize *over-generation* of questions by simplifying text sentences and utilizing this simplification as a source for question generation by applying transformations. This process produces a large set of candidate questions in which some may be deemed as unacceptable (due to conditions such as the vagueness of the generated question). The countermeasure applied in order to facilitate the discernment of which questions may be acceptable is a *statistical discriminate ranker*. Questions are manually evaluated and only a question that fulfills all the established requirements is deemed acceptable (27.3% of all questions). When the questions were ranked, top 20% included 52.3% of the acceptable questions which explains the usefulness of the ranking system. Also, ablation studies are made to figure how each of the utilized features affects the ranking system.

Dhole and Manning [16] introduce *Syn-QG*, a rule-based solution heavily dependent on the usage of syntactic rules that leverage the relationships between words in order to generate questions. This approach generates questions by identifying potential "short answers" using a combination of *five* independent strategies – dependency heuristics, Semantic Role Labeling (SRL) heuristics, generic entities, *VerbNet* [17] predicate templates, and *PropBank*[1] roleset specific natural language descriptions. After this process, their solution generates a set of question-answer pairs that are then back-translated (a process used to convert sentences into their "cleaner high probability counterparts"), providing the final result. Evaluation is performed using both BLEU [18] and a human process of evaluating the grammatical correctness and the relevance

[1] Available at https://propbank.github.io/

score of each provided question.

Khullar et al. [19] produce a syntax-based solution that automatically generates multiple questions using relative pronouns and relative adverbs. Their process consists of feeding input to a spaCy[2] parser where the system checks for the use of said pronouns and adverbs in the sentence. It looks for additional relationships between words in a sentence before the information is sent to their "rule sets" – a three step rule process that further explores sentences to retrieve the necessary information for the end result. They point out that there is no standard way to evaluate the output of a QG system, forcing the process to be evaluated manually. This procedure scores the system based on semantic adequacy, syntactic correctness, fluency, and distribution. One of the most important limitations pointed out in this research was that the system was meant to generate questions for sentences that contained at least one relative clause and this referred to only 20% of the used corpus.

Mitkov and Ha [20] utilize a NLP approach to generate Multiple Choice Questions (MCQs). This method combines transformational rules, automatic term extraction and a disambiguation mechanism. In this work, there is an emphasis on the structure of the sentences, having interest only in domain-specific Subject-Verb-Object (SVO) or Subject-Verb (SV) as these are the only transformable types. The created rules attempt to manipulate the SVO structure in order to generate acceptable questions.

The proposed approach by Omar et al. [13] is an automatic rule-based solution applied to the subject of computer programming using NLP techniques (such as the removal of stopwords and the usage of Part of Speech (POS) tagging) that support the classification of a given question in the context of BT by identifying important keywords and verbs.

2.2.2 Template-Based Approaches

Template-based approaches employ a defined structure which contains placeholders for variety and flexibility in handling different inputs.

Raynaud et al. [21] make use of a Knowledge Base (KB) created from hierarchically related topics which have resources assigned to them (wiki resources). This work, like the one by Mitkov and Ha [20], focus on the domain of MCQs. A question template is a tuple composed by three elements (S, P, R) where S refers to a set of stems that are organized using a dependency tree, P refers to a set of placeholders within S, and R refers to a sub-graph of the KB and is

[2] https://spacy.io/

responsible for providing the means to link placeholders to the correct answer. Existing QG templates are also reusable and previously used Question Answering (QA) templates are also convertible into QG templates which reduces the need to devise new QG templates.

Shirude et al. [22] create a domain-agnostic QG for structured data, specifically in the form of tables. This approach relies on the creation of templates and a “dynamic inbuilt tagger” — a mechanism that takes a data entry and assigns what they define as an entity. An entity recognizer is responsible for identifying and annotating the category of each column in the table based on the tagger’s dictionary. They claim that their system is scalable by allowing for the addition of new entities. Regarding their templates, it is stated that their “matching technique for templates is highly effective but doubts can be raised about its syntactical correctness and aptness at times”. This points out a common problem that is similar to Heilman and Smith [15] regarding question vagueness.

2.2.3 Neural Network-based Approaches

Researchers have more recently employed strategies to train NN that are able to create questions based on a textual input. Typically, NN approaches for QG interpret sentences as sequences of words (or even characters) that are converted into another sequence as output. These are known as Sequence-To-Sequence (Seq2Seq) approaches that implement what is called an *encoder--decoder* framework, often built using Recurrent Neural Networks (RNNs) such as LSTM or GRU (or, more recently, with the combination of several *attention mechanisms*, commonly known has transformers). In a Seq2Seq approach, the *encoder* takes the desired input and converts it into a set of hidden states that will be sent to the *decoder* for the decoding process, finalizing it by producing an output sequence which is then translated to an understandable output.

Zhou et al. [23] developed a Neural Question Generation framework which consists of a Seq2Seq [24] structure with a feature enriched encoder which contains the answer and lexical features such as Named Entity Recognition (NER), POS, and even the answer position to generate questions from natural language sentences. The solution employs an attention mechanism [25] to emulate the human process that is responsible for choosing the important content of a sentence. The model is built upon a Gated Recurrent Unit (GRU) [26] used in both directions of the input (from the beginning to end and in the reverse direction, also known as a BiGRU). The decoder also contains a GRU. To deal with unknown words, a copy mechanism is created so these words are essentially preserved from the input sentence without further applying any operations on them. This work relies on the Stanford Question Answering Dataset (SQuAD) [27, 28] dataset. Evaluation is performed automatically using the BLEU metric [18] and also manually by ranking questions using a scale from 1 (bad) to 3 (good). Results state that BLEU-4 score was 13.28 and the manual evaluation of the system is 2.18. *Precision* and *recall* are also calculated based on the WH-question type.

Bao et al. [29] research QG in a particular domain without labeled data by obtaining labeled data from other domains. The *doubAN* is a solution consisting of a generator and two discriminators, QA-Dis and DC-Dis, where QA-Dis stands for Question Answering Discriminator and its function is to provide more training data with estimated reward scores for generated text--question pairs, and DC-Dis refers to a Domain Classification Discriminator and its purpose is to help the generator learn the domain-general representations of input text. The generator uses an encoder-decoder framework that relies on RNNs. Their choice of this type of NN relates to the fact that each unit can be "modeled as a simple logistic sigmoid function and as complex as a GRU or a Long Short-Term Memory (LSTM)" [29]. Like Zhou et al. [23], a BiGRU is also present in this solution. Their QA-dis uses a reinforced learning algorithm to give a reward score to questions as feedback to their question generator in order to enhance the obtained results. Evaluation is performed using BLEU metric. Results show that the solution surpasses all tested baselines and BLEU is critiqued for not being able to subjectively evaluate results. The datasets used in this research are SQuAD (as unlabeled target dataset) and NewsQA [30] (as labeled source dataset).

Wang et al. [31] introduce *QG-Net*, a RNN solution for educational content based on information from educational textbooks. This is a Seq2Seq approach that uses a "reader -generator" (or encoder-decoder) framework. It encodes answers as part of the input and takes advantage of a pointer network [32] (which is a mechanism originally intended for the task of question summarization) in their question generation model with the purpose of creating output questions that are focused on specific parts of the provided input text. The *QG-Net* model is divided in two modules: *context reader* and *question generator*. The *context reader* uses a Bidirectional Long Short-Term Memory (BiLSTM) due to its capacity of "preserving past information in sequential learning tasks". The LSTM input is in fact built with Global Vectors (GloVe) [33] that are enriched using several linguistic features such as POS, NER and CAS (word case), through the process of concatenation. This module is also able to generate different questions for the same input based on the contextual information provided by answers. The *question generator* generates output in a word-by-word fashion where each word passes through an unidirectional LSTM and it is encoded into a fixed size vector, being then sent through a softmax function that calculates the final probability distribution as the addition of two probabilities: the probability distribution over the original question vocabulary and over the input context vocabulary, imposed by the pointer network. Results are evaluated by metric comparison with several other baselines. There is also a manual evaluation and a qualitative evaluation based on the display of generated questions. Results show that *QG-Net* outperforms on all metrics utilized (fluency, relevance, and preference), which aim to evaluate real-world applicability. Regarding datasets, *QG-Net* uses SQuAD to train the model which is then applied to OpenStax (OS)[3] textbooks for domain-specific analysis. Conclusions reinforce the idea that linguistic features are a very important part of the solution and that the proposed system is scalable. The only limitations pointed out are the lack of an absolute metric that can filter out questions (meaning expert review is necessary for maximum results and also that their solution is yet to be ready for real-world environments). Another significant limitation is the fact that it can only generate factual questions.

Liu et al. [34] introduce Clue Guided Copy Network for Question Generation (CGC-QG). The core is an *encoder-decoder* framework with both attention and copy mechanisms. Their proposed question generation model works by learning to identify where each word in the question comes from – either copied from the input or generated from a vocabulary. Also, given the answer, it attempts to figure what words can be copied from the input passage. A word is labeled as copied if it is a nonstop word shared by both the passage and the question and its word frequency is lower than a provided threshold. To help a neural model learn what to copy is a problem that is mitigated by predicting potential clue words (words that help reducing uncertainty of the model regarding how to ask a question or how to copy) in input passages. Words copied from input sentences to output questions are usually closely related to the answer chunk. Patterns of dependency are captured via Graph Convolutional Networks (GCN). To figure out which words are clue words, CGC-QG has a clue word predictor which utilizes a Syntatic Dependency Tree (SDT) which reveals relationships between answer token and other tokens in a given sentence. Based on the tree, a prediction of the distribution of clue words is made using GCN – encoder and a ST Gumbel-Softmax estimator for clue-word sampling. The outputs are then fed to an encoder as advisors of what may potentially be copied to the target question. Finally, the decoder

[3] available at https://openstax.org/

learns the probabilities of generating a word and copying a word from the passage. Copying is encouraged in this model. Results show that question words generated from the vocabulary are more frequent words while the not frequent ones are usually copied from input, which is normal in the human question generation process. Evaluation is performed using BLEU (precision), ROUGE-L (recall) and METEOR. The datasets used are SQuAD and NewsQA. Conclusively, it is pointed out that when removing extra feature embeddings such as POS, dependency types, and word frequency levels, the performance drops significantly. Without the clue prediction module, there is also a non-negligible performance drop.

2.3 Datasets

A dataset is the cornerstone of the training and testing process of any model. To adequately provide the best possible approach, a study on the currently used datasets is in bound. Section 2.3.1 reviews what are the most common datasets in the task of QG. Section 2.3.2 picks up on the difficulties in the current landscape of BT-oriented datasets for QC, and Section 2.3.3 takes into consideration all the previous information and justifies the choices that were made in this dissertation.

2.3.1 Datasets in Question Generation

The document presented by Amidei et al. [35] researches the evaluation methodologies used in a well defined six year span. Not only it reviews and organizes evaluation methodologies, but also displays information of datasets used. The Text-to-Text (Text2Text) approaches utilize a range of different datasets. In [35], it is stated that one of the problems that was found was the lack of *consistency* regarding the choice of datasets (complicating the process of comparing results), which means that converging to the same datasets is a positive behaviour that counterbalances this issue. The NN approaches presented rely mostly SQuAD or NewsQA or even both as the dataset of choice.

At the time of writing, two versions of SQuAD exist. The first version [27] is composed of over $100,000$ questions created by crowdworkers where the answer to any question is a segment of the respective text. It is a reading comprehension dataset that is large and widely used in the QG community. The second, and most recent, version of SQuAD is called Stanford Question Answering Dataset 2.0 (SQuAD 2.0) and is presented by Rajpurkar et al. [28]. SQuAD 2.0 brings increased difficulty by adding $50,000$ questions that are *unanswerable*, therefore forcing current NN models to adapt and develop mechanisms to understand when not to answer. Results show that the *F1-measure* value of a strong neural system (86%) could only attain (66%) for the same metric in SQuAD 2.0.

Similar to the characteristics of SQuAD, NewsQA [30] is a machine-comprehension dataset including over $100,000$ question-answer pairs based on a set of $10,000$ articles. This dataset is built on information collected through a meticulous four-staged validation process with the sole purpose of guaranteeing the quality of the dataset.

2.3.2 Datasets for the Revised Bloom's Taxonomy

When considering the task of QC, the required dataset must contain questions and respective labels according to RBT. We have identified that there are datasets that use RBT but with limitations in terms of *scope*, *availability*, *size*, and/or *consistency* [35].

We use the term *scope* to represent the requirement for a dataset to be as domain-agnostic as possible. We define the *availability* requirement it is expected that a dataset is publicly available at least for the research/educational purposes. *Size* is also key for a good classifier (even though most models for question classification do not require as much information as NNs models by resorting, for example, to manually-crafted heuristics). Finally, when considering *consistency* [35], it is expected that a dataset is used in several research works and by different authors in order to have some form of implicit agreement in its usage.

In the studied literature there are no agreed upon datasets contemplating the classification of RBT (or BT). To the best of our knowledge, there are no open-source, general-purpose datasets, in English, with enough information to reliably train a question classifier. Some of the studied literature uses as many as 100 entries that are divided into 6 categories, which provides little training for a classifier without any sort of rule-based approach.

An appropriate solution to this issue is to create a custom dataset. Although this does not solve the necessity of *consistency*, by customizing a dataset from publicly available information, it is our understanding that this dataset is the most suitable candidate for the requirements of this dissertation.

2.3.3 Final Regards Considering Datasets

To address the task of QG, most solutions work with SQuAD and/or NewsQA. Because there is such an implicit agreement and the fact that they present many similarities, this dissertation will proceed with the most commonly used dataset in the revised literature – SQuAD. However two problems remain: first, as mentioned by Wang et al. [31], solutions based on SQuAD are prone to generating factual questions (this is due do the fact that SQuAD is mostly comprised of factual questions). Secondly, there are no available datasets labeled with RBT.

To solve the first problem, the proposed approach is trained on SQuAD, providing a base model that can be *fine-tuned* with tailored questions for each of the taxonomy levels, resulting in multiple models specifically aimed to each level of RBT. This requires a second dataset that has higher diversity of questions. A dataset that follows the premisses of SQuAD and provides a higher diversity of questions is Question Answering in Context (QuAC) [36]. According to Choi et al. [36], QuAC's questions are often more open-ended (which can be translated into higher diversity and, consequently, be used to train a better model for each of the levels of RBT.

However, the second problem persists as QuAC is a large dataset that is unlabeled in terms of RBT. To solve this, we propose the creation of dataset that can be used to train a question classifier which can in turn be used to label larger datasets as well as work as a standalone classifier for other RBT-related applications. Therefore, and as previously hinted in Section 2.3.2, a custom dataset was created and is presented in Chapter 4.

2.4 Question Classification Overview

The purpose of QC is to attribute a tag or a label to a question according to its context using a set of specific guidelines. In the context of this dissertation, QC is necessary to label a dataset using the guidelines provided by the RBT.

2.4.1 Approaches in Question Classification and the Revised Bloom's Taxonomy

According to Silva et al. [37], a systematic literature review in the task of QC singled out a total of 80 studies. From these, 18.75% rely on Support Vector Machine (SVM) as the single algorithm in their solution while 15.0% use neural networks and only 10% focus solely on rules. When considering algorithm combinations, SVM is also the most common approach, appearing in 18 studies, followed by Naive Bayes (NB) which is presented in 11 studies. Rules are used in six studies. This research also reveals what taxonomies are most commonly used, presenting Li & Roth [38] in the first place with 26 studies followed by BT with 13 studies. Regarding metrics, this study shows that accuracy is the most used metric, far outweighing others and representing 68.75% of the studied pool. This paper also shows that most approaches that choose BT are often paired with rules while Li & Roth is often paired with neural networks. From all the considered studies, only *three* studies attempt to combine SVM with BT.

Yahya et al. [39] investigate the effectiveness of machine learning techniques such as SVM, NB, and K-Nearest Neighbours (k-NN) in the task of analyzing teacher's questions by classifying them using BT. These algorithms are evaluated individually and the chosen metrics for their evaluation are *accuracy* and *F1-measure* as well as sensitivity to the number of terms in a question. Results show the overall superior performance of SVM in all evaluated metrics.

Abduljabbar and Omar [40] propose a method to classify exam questions according to BT's *cognitive domain*. This method, used in programming questions, applies a combination of the classifiers studied by Yahya et al. [39]. In this study, each classifier is explored individually at first with and without feature selection methods like Chi-Square [41], Mutual Information, and Odd Ratio [40]. The classifiers are later integrated using a voting algorithm. In this work, the preprocessing steps are tokenization, stopword removal and the removal of all numbers and punctuation marks. The selected evaluation metric was *F1-measure*. Although results show that SVM and k-NN already provide similar results throughout the six domains of BT, it is also stated that the voting approach surpasses any individual algorithm by providing the highest scores.

Osadi et al. [42] also provide an ensemble solution combining rules with SVM, NB, and k-NN, testing each individual model with a dataset of 100 programming examination questions. The ensemble was able to produce an accuracy that exceeded the value achieved by the SVM solution. This study points out the need to craft many rules and its consequent difficulty in terms of scaling. This work also points out the need to deal with ambiguous questions by creating a combination module using *WordNet* similarity and majority voting.

Osman and Yahya [43] use a dataset with 600 exam questions of an English language course to test several classification models. The classification itself is motivated by linguistic features such as the Bag-of-Words (BOW), POS, and N-grams.

Other studies, such as the work of Kusuma et al. [3], attempt to classify questions using SVM with questions written in Indonesian. They use a dataset containing 130 questions that is split into their test and training data. Their proposed solution preprocesses information in four steps — tokenization, stopword removal, stemming, and POS tagging. This information is then fed to a process of feature extraction which has the purpose of finding additional information (lexical and syntactical) that may be relevant for the classification process.

The proposed approach by Omar et al. [13] is an automatic rule-based solution applied to the subject of computer programming using NLP techniques (such as the removal of stopwords and the usage of POS tagging) that support the classification of a given question in the context of BT by identifying important keywords and verbs.

Haris and Omar [11] show the usage of a hybrid approach that combines rule-based mechanics with a statistical classifier. A set of 64 rules is created, encompassing all *six* levels of BT. The statistical method is essentially a N-gram classifier. This work acknowledges that a rule-based method helps determining question categories and that it does provide good results in certain situations. However, it is also stated that this method “can be quite tedious as it can be time-consuming”. Finally, they praise their hybrid solution as this combination attempts to mitigate the individual weaknesses of both rule-based methods and N-grams. As evaluation metric, *F1-measure* was selected and, according to the results, the solution provides an F1-measure that is 1.34 times better than their rule-based experiment and 1.7 times better than their N-grams experiment, which confirms their initial expectations of having a hybrid approach that can surpass each mechanism individually.

3 Question Classification

In this chapter, we provide an in-depth presentation of the question classifier as well as its constituents.

This chapter is organized as follows: Section 3.1 introduces the selected approach and the expected challenges posed by this task. Section 3.2 provides information on the dataset as well as its distribution; Section 3.3 details a comprehensive plan of all the steps involved in the classification process; Section 3.4 provides an overview of the experimental setup used; finally, Section 3.5 provides the relevant results from this portion of the dissertation.

3.1 Approach and Challenges

In the scope of this dissertation, the need to have a question classifier stems from the fact that there are no agreed upon question datasets based on Revised Bloom's Taxonomy (RBT) (or Bloom's Taxonomy (BT)). Instead of manually labeling a dataset to be used directly in the process of question generation, the process of creating a suitable dataset can be simplfied by developing classifier that can label existing datasets as an attempt to adapt them to the context of RBT.

The task of classification with this taxonomy in particular poses a challenge regarding how to deal with sentences that could arguably belong to more than one class (ambiguous cases). Naturally, another challenge is to obtain an initial dataset that can provide enough insight for an accurate classifier.

As previously mentioned, the purpose of Question Classification (QC) is to classify a dataset that can be used to train a Question Generation (QG) model that respects the premises of RBT. Although the classifier exists to solve the problem of obtaining acceptable datasets, it can hold even more importance by working as a reinforcement classifier to label the questions generated by the QG, allowing for a better understanding of the outputs of the QG model.

3.2 Question Classification Dataset Overview

As previously mentioned, the lack of available datasets annotated with BT is a current limitation. To overcome that, we created our own dataset. It was created using questions from sources such as OpenStax (OS) [31] and a set of RBT-related sites[4]. These questions were then labeled

[4] Available at https://bloomstaxonomy.org/ and https://www.mandela.ac.za/

according to their respective *cognitive level*. The final dataset is comprised of 642 questions manually labeled by a single person, fluent in English (non-native) according to the indications provided by Bloom et al. [4] and Krathwohl [5]. The distribution of the dataset by each one of the taxonomy levels is shown in Table 3.1.

As shown in Table 3.1, the dataset has an unbalanced distribution which is a consequence of the added difficulty of devising questions of higher levels in the taxonomy. To further complement the information on Table 3.1, Table 3.2 provides an example of dataset entries.

3.3 Model Overview

In order to create the model, a training pipeline was built by combining *preprocessing*, *feature extraction* and *resampling* techniques, which are then applied to the dataset, converting it into a suitable source of input for the training process. Once trained, the model that yields best results will be considered the final classifier.

Table 3.1: Distribution of the dataset used for QC.

Cognitive Level	Questions
Remember	122
Understand	135
Apply	121
Analyze	145
Evaluate	72
Create	51

Figure 3.1 provides a visualization of how the different techniques were applied to the data, showing how all the models were trained in the classification process.

3.3.1 Preprocessing

By *preprocessing* the input, it is expected that the number of tokens to work with is reduced, simplifying the work necessary to produce a question classifier (and enhancing its results). All of the information was converted to lowercase and all the numerical values were replaced with the `<NUM>` token. All the information was then stripped of its punctuation. The reason behind this is that, despite we are classifying questions, many of the questions in the dataset are written in the

Table 3.2: Example entries of the dataset used in QC.

Question	Cognitive Level
Why are ionized gasses typically found in very high-temperature environments?	Understand
How does activity on the Sun affect natural phenomena on Earth?	Analyze
What is your attitude toward mental health treatment?	Evaluate
Describe an event schema that you would notice at a sporting event	Understand

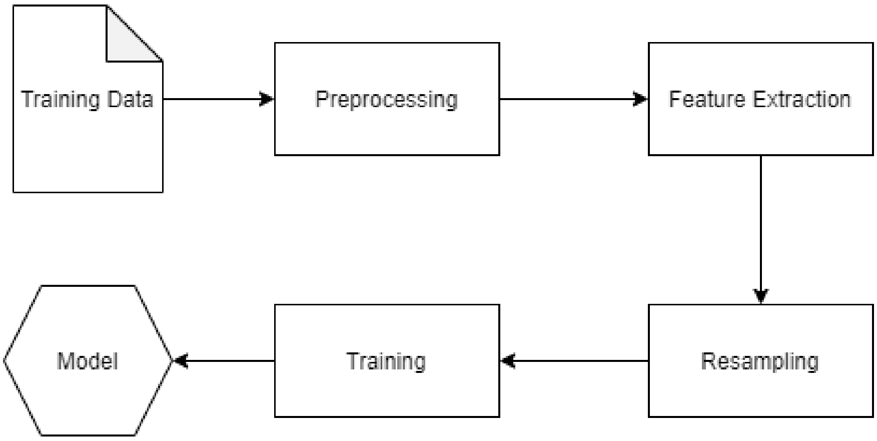


Figure 3.1: Classifier training workflow.

imperative form rather than the *interrogative* form. Finally, all the information is *lemmatized* using *SpaCy* [44] to further simplify the input text.

3.3.2 Feature Extraction

Feature extraction is performed using with two strategies: *CountVectorizer* and *TfIdfVectorizer*. These strategies differ in how the word frequency is interpreted. While *CountVectorizer* interprets a collection of documents as a matrix of regular word occurrences, *TfIdfVectorizer* builds upon the former by building a matrix of *TD-IDF* word weights. These methods were selected as they facilitated the implementation of complementary strategies such as a *N-gram* feature extraction mechanism and a word *tokenizer*, as well as the definition of the minimal number of occurrences of a word in the dataset — which is set to the top 90% words.

3.3.3 Resampling

Resampling refers to a combination of *oversampling* and *undersampling* and it is a mechanism that is used when the amount of available data needs to be altered (usually when the dataset lacks information or is unbalanced).

The created dataset, although directed towards general purpose applications, suffers from being unbalanced. By oversampling, artificial information is created as an attempt to balance the dataset. Strategies of oversampling may vary depending on the objective as it is possible to target specific classes or specific groups of classes. In this dissertation, a combination of *Synthetic Minority Oversampling Technique (SMOTE)* [45] and *RandomUnderSampler* are selected, with a configuration of *not majority* and *all*, respectively. The *not majority* configuration allows for the resampling of all other classes except the class with most entries in the dataset — *Analyze*.

3.3.4 Model Selection

Based on what was gathered from the literature review, it was to be expected that Support Vector Machine (SVM) would outperform other models. In our process, we have included SVM [46, 47], Naive Bayes (NB) [48], Logistic Regression (LR) [49, 47], K-Nearest Neighbours (k-NN) [50], GradientBoosting (GB) [51, 52] and AdaBoost (AB) [53]. All the presented implementations were obtained using *Scikit-Learn (Sklearn)*[5].

Support Vector Machines In this dissertation, the implementation of SVM used is the *LinearSVC*, which is, according to documentation, more flexible than a regular SVC with a linear kernel. This algorithm implements a *One-vs-Rest* approach, which can be interpreted as breaking down a multi-class problem into several binary problems. Preliminary tests with both LinearSVC and SVC with a linear kernel show that *LinearSVC* is faster to produce slightly better results.

Naive Bayes Two algorithms were selected: *Multinomial Naive Bayes (MNB)* and *Complement Naive Bayes (CNB) [54]*. The former is, according to Sklearn's documentation, a suitable solution for text classification based on word counts and is also often an adequate solution for *TD-IDF* counts. According to Rennie et al. [54], CNB was created with the purpose of correcting the "severe assumptions"made by the former (regarding the selection of weights for the

[5] Available in https://scikit-learn.org/stable/modules/classes.html

decision boundary and the assumption that features selected by the former algorithm are independent). In the context of this dissertation, the two algorithms are compared to understand which provides better results.

Logistic Regression This statistical model typically works by applying a logistic function (also known as a sigmoid curve) to solve binary problems. Regarding multi-class implementations, the Sklearn's version of this model is different than the *LinearSVC One-vs-Rest* approach. It instead uses a "true multinomial logistic regression model" which is explained by Bishop [55] and in order to be able to use it, we rely on the default solver which is the Limited-Broyden–Fletcher–Goldfarb–Shanno (lbfgs) [49].

K-Nearest Neighbors Sklearn's documentation for k-NN references the work done by Stevens et al. [56] as the basis of the implementation, where the classification is computed from a majority vote of the nearest neighbors of each point. In order to select a neighbour, the implemented approach resorts to distance calculated using the *minkowski* metric. The selected value of neighbors is the same as the number of categories — six.

GradientBoosting Unlike the previous methods, *GradientBoosting* is an *ensemble* algorithm. It is an additive model that works in a stage-wise fashion (iteratively). The principle behind *GradientBoosting* is to optimize models by working with a set of weak prediction models (often decision trees), called *weak learners*, that are sequentially being improved to provide a better result. In this particular implementation, the ensemble creates as many regression trees as the number of categories in the problem and attempts to optimize the result by iteratively selecting hypothesis that points in the negative gradient direction.

AdaBoost *AdaBoost* is an *ensemble* algorithm. This implementation recreates the *Multi-class AdaBoost* presented by Zou et al. [57]. This *ensemble* classifier works with several copies of a specific classifier (in the implemented scenario, a *Decision Tree* classifier) on the same dataset. Beginning with a *weak learner*, the principle behind this is optimization by adjusting the weights of previously incorrectly classified entries so that the following classifiers focus on resolving these entries, providing an enhanced result.

3.4 Experimental Setup

For the task of classification, several test configurations were created in order to thoroughly evaluate which is the most suitable algorithm for the proposed classifier. The tests were performed twice, varying the selected feature extraction method (*CountVectorizer* and *TextVectorizer*), as this is a necessary step to ensure that the information flows correctly throughout the training pipeline. The tested scenarios are presented in the following list:

- *baseline* – This is first test where the input is yet to be preprocessed and is directly sent to the feature extraction module.
- *pre* – Building upon the baseline, the results shown in this test include the preprocessing steps (excluding lemmatization).
- *lemma* – Building upon the *pre* configuration, adding the lemmatization process.
- *ngrams(1,2)* – Building upon the *lemma* configuration, this experiment adds unigrams and bigrams using the implementation specified in Section 3.3.2.
- *ngrams(1,3)* – Building upon the *lemma* configuration, this experiment adds trigrams to the previous item, following the implementation specified in Section 3.3.2
- *ngrams(1,2) + smote* – This configuration drops the trigrams, using only unigrams and bigrams. It also adds SMOTE oversampling technique as mentioned in Section 3.3.3.
- *ngrams(1,2) + smote-ru* – In this final configuration, SMOTE is complemented by the undersampler strategy defined in Section 3.3.3.

3.5 Results

Tables 3.3 and 3.4 present the accuracy values obtained both in the training set (with cross-validation) and in the test set, respectively. Both tables display results for the all the different models with all the mentioned settings in Section 3.4. In an attempt to further enhance the accuracy of the proposed classifier, we present our *fine-tuning* approach in Table 3.5. Using the custom dataset mentioned in this Chapter, we divide the data in a $80/20$ split of training and testing respectively. This was performed using Sklearn's *train_test_split*.

Table 3.3: Classifier cross-validation accuracy for all the tests performed.

Cross-Validation	SVM	MNB	CNB	LR	k-NN	AB	GB
TfIdfVectorizer							
baseline	0.597	0.513	0.556	0.571	0.521	0.363	0.567
+pre	0.585	0.536	0.538	0.569	0.551	**0.440**	0.581
+lemma	0.588	0.506	0.518	0.559	0.495	0.419	0.579
+ngrams(1,2)	0.608	0.528	0.587	0.559	**0.573**	0.429	0.588
+ngrams(1,3)	0.596	0.502	0.587	0.526	0.571	0.403	0.591
ngrams(1,2)+ Smote	0.784	**0.747**	**0.748**	0.774	0.423	0.386	0.708
ngrams(1,2)+ Smote-ru	**0.803**	**0.747**	0.745	**0.788**	0.415	0.416	**0.739**
CountVectorizer							
baseline	0.616	0.547	0.536	0.583	0.347	0.329	0.587
+pre	0.616	0.572	0.538	0.602	0.387	0.435	0.596
+lemma	0.595	0.539	0.530	0.608	0.409	0.442	0.571
+ngrams(1,2)	0.612	0.592	0.551	0.612	0.345	0.447	0.59
+ngrams(1,3)	0.617	0.598	0.538	0.606	0.288	0.427	0.589
ngrams(1,2) + Smote	**0.642**	0.624	**0.606**	0.655	0.463	0.505	**0.671**
ngrams(1,2) + Smote-ru	0.639	**0.628**	**0.606**	**0.662**	**0.464**	**0.508**	0.661

In Table 3.3 it is possible to confirm that, in all cases, the defined training pipeline allows for much better accuracy values than the baseline. It is also possible to confirm that SVM outperforms all other models in terms of accuracy with a value of 80.3%. The cross-validation setup suggests that the *preprocessing* and the *lemmatization* processes negatively affect the outcome of the final SVM model. However, ablation tests were performed by removing these steps and the final result was 2.1% less accurate (achieving a cross-validation score of 78.2%) than what is presented in Table 3.3. This table also exhibits that although SMOTE benefits both *CountVectorizer* and *TextVectorizer*, the results are highly accentuated when using the latter. This suggests that the *resampling* process functions better with *TF-IDF* weights rather than raw counts.

As SVM was the best performing model, we studied if any improvements were possible. SVM allows for fine-tuning through the adjustment of its C hyperparameter. It functions as a regularization parameter that adjusts the separation hyperplane between categories. Altering it informs the algorithm of how important it is to correctly classify each instance (increasing C)

Table 3.4: Classifier test accuracy for all the tests performed.

Test	SVM	MNB	CNB	LR	k-NN	AB	GB
TfIdfVectorizer							
baseline	0.643	0.535	0.582	0.581	0.496	**0.411**	0.527
+ pre	**0.667**	0.581	0.589	0.605	0.550	0.357	0.527
+ lemma	**0.667**	0.550	0.558	0.597	0.566	0.389	0.550
+ ngrams(1,2)	0.659	0.542	0.643	0.612	**0.581**	0.286	0.511
+ ngrams(1,3)	0.659	0.542	**0.651**	0.582	0.566	0.287	0.519
ngrams(1,2) + Smote	**0.667**	0.604	0.573	0.627	0.333	0.295	0.558
ngrams(1,2) + Smote-ru	**0.667**	**0.6511**	0.612	**0.643**	0.379	0.302	**0.558**
CountVectorizer							
baseline	0.627	0.597	0.535	0.636	0.295	0.295	0.504
+ pre	0.659	0.628	0.566	0.674	0.395	**0.457**	0.543
+ lemma	0.651	0.604	0.604	0.628	**0.411**	0.357	0.519
+ ngrams(1,2)	**0.667**	0.620	0.627	**0.674**	0.295	0.302	0.512
+ ngrams(1,3)	**0.667**	0.604	**0.643**	0.6511	0.248	0.302	0.503
ngrams(1,2) + Smote	0.604	**0.667**	0.612	0.573	0.3488	0.426	0.542
ngrams(1,2) + Smote-ru	0.604	**0.667**	0.612	0.574	0.37	0.426	**0.573**

rather than finding an hyperplane that has a largest minimum margin (decreasing C). There are no specific steps to adjust it, as each problem requires specific testing. Table 3.5 presents a study performed on the same dataset when adjusting parameter C using a fixed set of values presented by Burkhart and Shan [58]. An increase of 0.4% was obtained by increasing the C hyperparameter from the standard value (10^0) to 10^2. It is also visible that reducing this value results in a lower accuracy. To the best of our knowledge, this implementation of SVM does not have other ways of fine-tuning. The final accuracy value of the proposed classifier, for the train set, is then 80.74%.

The defined pipeline also works in most cases, as it is possible to see in Table 3.4. The exception to this result is the case of *AB*, where the provided configurations worked poorly. The best configuration for AB was still worse in terms of accuracy than the baseline model for SVM. In Table 3.4, SVM is also the best model with an accuracy of $66.7.\%$. Another relevant result that the CNB performs worse than MNB in our experimental setup, contrary to what was

Table 3.5: Cross-validation accuracy fine-tuning using SVM C hyperparameter.

C	Cross-validation accuracy
10^{-1}	0.785
10^{0}	0.803
10^{1}	0.805
10^{2}	**0.807**

expected based on Sklearn's documentation.

It is then confirmed that SVM is, as expected from the literature review, the most suitable algorithm to be used for the task of QC using RBT.

To provide further context on the obtained classified entries, Table 3.6 displays questions that were classified using the classifier.

Table 3.6: Example entries of the dataset used in QC.

Question	Cognitive Level
What is the real name of WWE Superstar Edge ?	Remember
How would you explain the experiment of Schrödinger's cat ?	Understand
Calculate the temperate of water at surface level in Celsius	Apply
Compare and contrast the advantages and disadvantages of current generation consoles.	Analyze
Evaluate the Bill of Rights and determine which is the least necessary for a free society.	Evaluate
Develop a new application for smartphones that allows for the detection of Covid-19 patients.	Create

4 Question Generation

In this chapter, a comprehensive explanation of the generation process is made. Section 4.1 puts in perspective the considerations made as well as the challenges presented by the task of Question Generation (QG). Section 4.2 displays an overview of the question generation model, providing an explanation of how the *encoder* is developed, as well as the *decoder* and the complementary *attention mechanism*. Section 4.3 details the overall experimental setup, including the training process and the evaluation procedure employed. Finally, Section **??** presents the relevant results regarding the question generator, reporting both automatic and manual evaluation procedures.

4.1 Considerations regarding QG

Considering all the literature explored in Section 2.2, some remarks are essential in the process of selecting a technique: Wang et al. [31] state that prior research regarding rule-based models [15, 59] works reasonably well because it has "meticulously crafted heuristics" with well structured input. However, if features in the input are not predictable in the rules, the output will not be satisfactory. Furthermore, it also states a paradigm shift from rule-based models to data-driven models as is the case of Recurrent Neural Networks (RNNs) which are present by Wang et al. [31] and Zhou et al. [23], and other Sequence-To-Sequence (Seq2Seq) approaches. Another commendable factor in this research is the fact that it surpassed the state-of-the-art rule-based approach presented by Heilman and Smith [15].

Liu et al. [34] also defend that heuristic approaches such as rules and templates rely on manually created heuristics that require expert work in their conception and therefore are expensive and not scalable or diverse. On the other hand, neural-network based approaches are used precisely because they lack this need of expert in their conception (but, as mentioned by Wang et al. [31], expert review is still a factor to consider given no evaluation metrics perform at equivalent level).

Bao et al. [29] provide a similar opinion defending that the human effort necessary in rule and template-based approaches can be highly reduced by the usage of neural network models such as Seq2Seq with an *encoder-decoder* framework.

The proposed solution in this dissertation intends to provide scalabilty, adaptability and diversity with minimal expert contribution. As discussed in the previous subsections, Neural Network (NN) fit the necessary conditions by allowing models to be trained in conjunction with several Natural Language Processing (NLP) processes and without the need to create any specific heuristics.

One of the most important challenges is to generate questions disregarding the answer as part of the training process. The natural process of question formulation question comes from the need to know the answer to a specific information. Also, by removing the need to have every question answered, future dataset creation for question generation can be simplified.

The biggest challenge is to work with the commonly used QG datasets. Most approaches work with datasets that lead to the generation of factual questions as stated by Wang et al. [31]. When considering the necessity of respecting Revised Bloom's Taxonomy (RBT), purely factual questions are not diverse and often fall in the *remember* and *understand* cognitive levels, leaving out the remaining *four* categories. A final obstacle is to overcome the problems regarding question vagueness and acceptance [15] in an attempt to produce questions that are indeed valuable for a learning environment.

4.2 Model Overview

The proposed QG solution is a *Seq2Seq* model. The general principle behind these models is to accept as input a sequence of items (words, characters or even images) and provide another sequence as an output. This is achieved by implementing what is called an *encoder-decoder* framework where two RNN are paired in order to accomplish the desired result (this can also be performed with the use of *transformers* [25]). The *encoder* is responsible for taking the input and encoding it into a set of hidden states which will then be sent to the *decoder* with the objective of translating the encoded content, which will ultimately be converted into understandable output. Figure 4.1 depicts a detailed view of the architecture proposed. The following sections describe the *encoder*, *attention mechanism*, and *decoder*.

4.2.1 Encoder

As mentioned in Section 4.2, the *encoder* is responsible for obtaining the input and applying the necessary steps before it transforms it into a set of hidden states that will be sent to the corresponding *decoder*. Before being sent to the *encoder*, the input is *preprocessed* by being converted to lowercase and removing any trailing whitespaces. This input is then sent through an embedding layer, which has been initialized using Global Vectors (GloVe) [33]. Once the input is processed into vector representations, it is sent to the *encoder*.

The encoder uses a Bidirectional Long Short-Term Memory (BiLSTM) [60, 61], due to its ability of “preserving previous information in sequential tasks” [31]. Once in the encoder, the input will be converted into information that will be processed by the decoder.

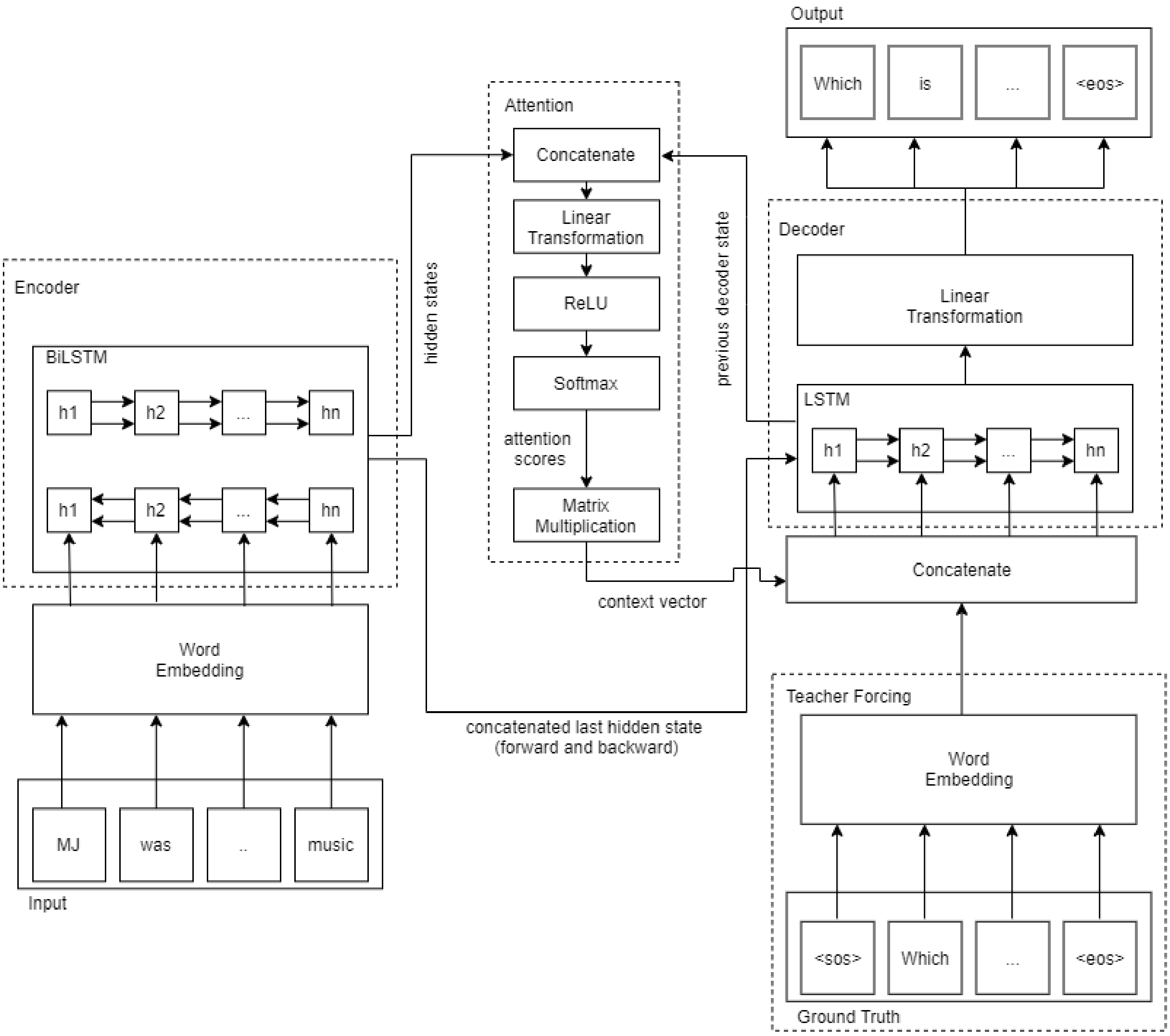


Figure 4.1: Seq2Seq model detailed representation. The exhibited model presents the setup utilized to train the base-model for question generation.

In the proposed solution, the *encoder* produces two outputs – a set of all LSTM states and a concatenation of the rightmost hidden states in both directions of the LSTM. The former is sent to the *attention mechanism*, explained in Section 4.2.2, and the latter is sent directly to the initial state of the *decoder*, a process that is discussed in Section 4.2.3.

4.2.2 Attention Mechanism

The *encoder* output states can be directly sent to the *decoder*. However, results can be enhanced by using an *attention mechanism* [62]. The purpose of this mechanism is to emulate the way humans read and process information, which essentially translates to focusing on specific keywords rather than reading the complete text from left to right. By receiving the output from the *encoder* and the previous hidden state from the *decoder* the *attention mechanism* returns a *context vector* that will be then fed to the *decoder*.

In the proposed approach, the inputs of the *attention mechanism* are concatenated and sent through a linear transformation layer. The output of this operation is then sent to a Rectifier Linear Unit (ReLU) activation function. Once calculated, this is sent to a *softmax* function which has the purpose of computing *attention scores* that will be used in the calculation of the *context vector*. The context vector is essentially a matrix multiplication of the calculated *attention scores* and the *encoder* hidden states. Once computed, the *context vector* is sent to the *decoder*.

4.2.3 Decoder

The purpose of the *decoder* is to convert the BiLSTM states obtained from the *encoder* and, using a single LSTM, convert them into an output sequence. The design of the decoder is displayed in Figure 4.1. When inferring, the input used in the decoder comes from two sources: the last state obtained directly from the encoder and the *context vector* generated by the *attention mechanism*. When training, the second input changes — it becomes a concatenation between the *context vectors* obtained by the *attention mechanism* and the GloVe embeddings obtained from a concept known as Teacher's Forcing [63].

Teacher's forcing is employed with the goal of "teaching" the NN by feeding the decoder inputs from the training dataset (*ground truth*) rather than using what the network outputs in a given time step. This way, it is expected that the LSTM converges faster and provides better results. The words that were forced also pass through an embedding layer using GloVe embeddings. Once converted into embedding vectors, these are concatenated with the context vector obtained from the *attention mechanism* and this result is then the input of the decoder LSTM. This LSTM has its initial state set to the concatenation of both directions of last hidden state of the *encoder* and is then ready to process the information. Once decoded, the information goes through a *linear transformation* layer, producing an array of predictions with the probability of each word. From these predictions, the highest value is assumed to be the correct word and is then converted to its textual representation using a dictionary with all the words provided in the training dataset.

4.3 Experimental Setup

4.3.1 Training and Fine-tuning

Once defined, the Seq2Seq model is trained using the complete Stanford Question Answering Dataset (SQuAD) training set. When training this model, several parameters have to be managed in order to produce the best results possible. The proposed training configuration consists of a duration of 14 epochs, using a BiLSTM with a *hidden size* of 256 in the *encoder* and a single LSTM with a hidden size of 512 in the *decoder*, a *batch size* of 64 and an embedding size of 300. Regarding loss, the selected loss function in this approach is *Cross Entropy*. For the optimization strategy, a preliminary study was made comparing Stochastic Gradient Descent (SGD), RMSprop (RMSprop), and Adaption Moment Estimation (Adam), with the goal of selecting the most suitable optimizer for the proposed approach. In this study, all models were trained for the same amount of epochs and the one that produced the lowest loss values was the one using Adam.

The training yields a base model which is still not final. Instead, we *fine-tune* it in six different iterations, generating one model for each level of the RBT, using portions of the Question Answering in Context (QuAC) dataset that were previously selected by classifying the whole dataset with the approach developed in Chapter 3, and then splitting the dataset based on the classified questions.

All the models are trained and tested using an $80/20$ split. The *fine-tuning* process shares the same optimization and loss function as the training process but it is trained for 10 epochs with a batch size of 16 (preserving all other training configurations). A visual representation of the processes of training and *fine-tuning* is shown in Figure 4.2.

4.3.2 Evaluation Methodology

We evaluate the proposed process of QG using an automatic and a manual approach. The reason why both styles of evaluation are performed is because there are no automatic metrics that can capture the subjective matters regarding question evaluation, as discussed by Khullar et al. [19],

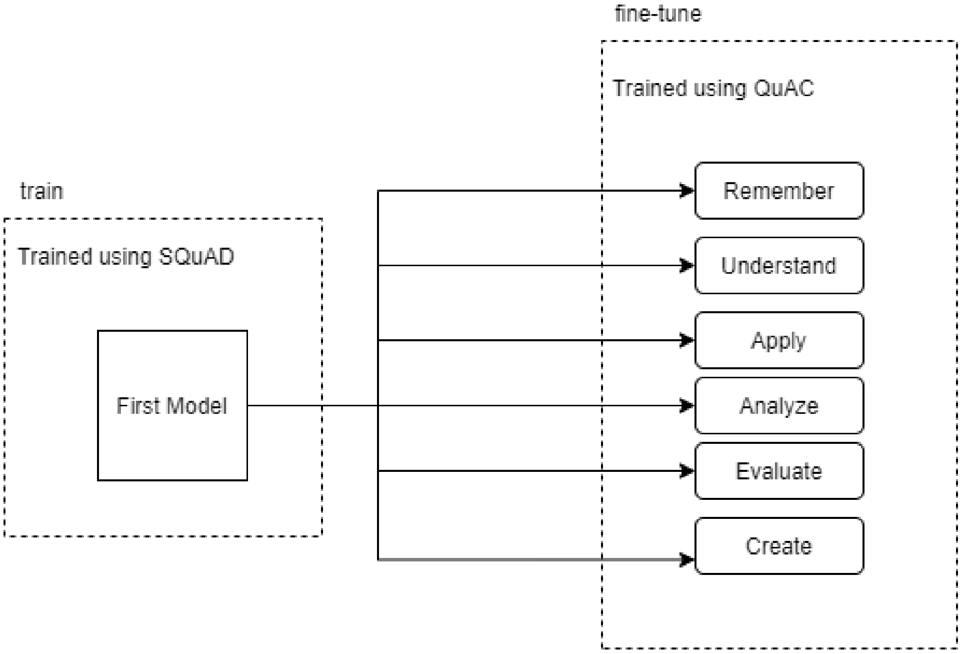


Figure 4.2: Generated models in the training and fine-tuning processes.

Wang et al. [31] and Bao et al. [29]. *Automatic evaluation* is performed by calculating the BLEU-1 [18] as it is a commonly used metric in Seq2Seq tasks and employed in multiple similar evaluation processes [23, 29, 34]. We use it to perceive how the fine-tuned models vary in terms of this metric. Furthermore, the generated questions are also classified using the classifier defined in Chapter 3, to understand if each one of the fine-tuned models produces a question of its respective taxonomy level.

Despite the selected automatic approach, one of the biggest issues in QG is the lack of a metric that can determine the quality of a solution as current metrics fail to grasp subjective concepts such as context. Because of this, researchers often prefer to either complement the automatic evaluation with a manual counterpart or even only perform customized manual evaluations.

We propose a manual evaluation that attempts to capture four parameters of a question:

- *Well Written* – it attempts to understand if a question is grammatically *well written*, evaluated on a scale of 1 to 3, with 1 being poorly written and 3 being well written.

- *Understandability* – it asserts if it is possible to know the purpose of the question despite how it is written, evaluated on a scale of 1 to 3, with 1 being not possible to understand and 3 being understandable.

- *Relevance* – it quantifies how relevant the question is for the provided context, evaluated on a scale of 1 to 3, where 1 is attributed to irrelevant questions and 3 is given to highly relevant questions.

- *Answerability* – it verifies if an answer to this question is present in the text, evaluated binarily with a *yes* or *no*.

4.4 Results

4.4.1 Automatic Evaluation

For all of the fine-tuned models, the test set is the remaining 20% from the 80/20 splits. For the base model, the BLEU-1 score was calculated using the SQuAD test set, attaining a result of 9.81. Table 4.1 displays all the obtained BLEU-1 scores for each of the fine-tuned models. Results show that there is a variation of BLEU-1 scores between the different levels of the taxonomy, with the highest BLEU score being achieved by the *apply* model with a value of 15.53, and the lowest score being obtained by the highest level of RBT – *create* –, with a value of 9.86.

To further inspect the quality of the results and in order to understand how often each fine-tuned model produces questions of its respective taxonomy level, we classified them with our Question Classification (QC) described in Chapter 3. The percentage of generated questions, for each level, that match the expected taxonomy level can be seen in Table 4.1.

From this analysis it is shown that the *fine-tuning* process facilitates the generation of questions of a specific RBT level, having all models produce questions of their respective level over 67% of the time. Upon further inspection to the variety of questions produced by each model, we confirmed that all models produce different questions for different contexts with the exception of the *create* model. The reason behind this is that the information used to trained this model is extremely biased into a very restricted set of questions, hindering the generation process.

4.4.2 Manual Evaluation

Using these parameters explained in Section 4.3, a survey was created containing 37 questions, from which 4 were control questions consisting in two positive examples and two negative examples. The remaining 33 examples were selected at random but using a balanced distribution of questions per taxonomy level (except the create level due to the bias explained in the last section), as shown in Table 4.2. The created survey was then filled by 8 evaluators with different

Table 4.1: BLEU-1 metric scores for each of the fine-tuned models, along with the classification accuracy of the generated questions according to the expected taxonomy level.

	Remember	Understand	Apply	Analyze	Evaluate	Create
BLEU-1	12.23	12.57	15.53	13.23	13.88	9.86
Accuracy (%)	85.90	75.60	86.00	78.00	67.50	90.00

Table 4.2: Distribution of the survey questions by level of the revised Bloom's Taxonomy.

	Remember	Understand	Apply	Analyze	Evaluate	Create
# questions	6	6	6	6	6	3

educational backgrounds. As an elimination criteria, the level of proficiency in English from a scale of 1 to 5 was questioned. All surveys that containing an answer lower than 3 were to be discarded, but all evaluators reported a score of 3 or higher. The complete questionnaire is presented in Appendix A.

To analyze the results, we calculated the average score of the submitted surveys regarding each of the evaluated parameters. Since *answerability* works in a different scale, those results are showing the percentage of positive answers. To further validate our results, we calculate *inter-rater agreement* using Fleiss' Kappa [64]. To interpret the values of Fleiss' Kappa, we follow the guidelines proposed by Landis and Koch [65]. We also computed the accuracy of the models in generating questions of the appropriate type, measured by our own classifier.

Tables 4.3 and 4.4 show the evaluated metrics regarding the four control questions. Analysing the results, there is a clearly defined separation between the results on positive and negative control questions. A positive control question refers to a question that should attain the highest value possible in a given parameter (3/*yes*), while the negative control question should achieve the lowest value possible (1/*no*). Attending to this description, it is clear that the control questions have average scores and answerability percentages that respect the intended purpose. When considering Table 4.4, it shows that the *inter-rater agreement* is much higher in classifying the *well written* parameter rather than *understandabilty* or *relevance*. This happens due do the contrast between the objectivity of the first parameter and the subjectivity of the remaining parameters. A well written question follows a specific set of tangible rules that allow evaluators to understand if a question is in fact well-written while the remaining parameters are highly subject to

Table 4.3: Average scores obtained on the control questions, for all parameters. Scores for the first three parameters range between 1 and 3, where 3 is the best score possible. Answerability is a percentage of positive answers.

	Well Written	Understandability	Relevance	Answerability (%)
Positive Control Qs	2.84 ±0.37	2.69 ±0.44	2.69 ±0.44	81.30
Negative Control Qs	1.13 ±0.23	1.38 ±0.83	1.31 ±0.60	12.50

Table 4.4: Fleiss' Kappa and Agreement percentage for the control questions.

Positive	Well Written	Understandability	Relevance
Fleiss Kappa	0.65	0.38	0.38
Agreement %	76.79	58.93	58.93
Negative	Well Written	Understandability	Relevance
Fleiss Kappa	0.68	0.33	0.49
Agreement %	78.57	55.36	66.07

interpretation. The highest and the lowest values of *inter-rater agreement* come from the negative control questions (the highest being the *well written* parameter with a value of 0.68 and the lowest being *understandability* with a value of 0.33).

Considering the 33 non–control questions, the average scores are presented in Table 4.5, where it is possible to confirm that all values exceed the lower half of the scale, with the lowest value belonging to *relevance* (2.13), and the highest value being the *well written* parameter (2.39). When considering the *answerability* percentage, roughly half the questions are unanswerable with the provided context. The reason why this is likely to happen is due to the fact that the generation process does not use the answer tokens as part of the generation process, therefore allowing it to be less specific. Considering Fleiss' Kappa and the agreement percentages, shown in Table 4.6, the highest agreement value belongs to the *well written* parameter (0.35), while the lowest is obtained by the *understandability* parameter (0.21). However, and as expected, these values are significantly lower than the ones shown in the control scenario, having the *understandability* parameter barely reach the classification of*fair agreement* according to Landis and Koch [65] From this we can derive the difficulty of the .

Table 4.5: Overall average scores obtained on all questions, for all parameters. Scores for the first three parameters range between 1 and 3, where 3 is the best score possible. Answerability is a percentage of positive answers.

Well Written	Understandability	Relevance	Answerability (%)
2.39 ±0.54	2.33 ±0.73	2.13 ±0.67	50.76

Table 4.6: Fleiss Kappa and Agreement percentages on all questions, for all parameters.

	Well Written	Understandability	Relevance
Fleiss Kappa	0.35	0.21	0.23
Agreement %	56.71	47.51	48.38

When organizing the questions in their RBT cognitive level, it is possible to analyze how each of the fine-tuned models behaves in terms of the studied parameters. Tables 4.7 and 4.8 detail the results by the six taxonomy levels. Considering the *well written* parameter, it is visible in Table 4.7 that all values are fairly above 2.0, with the exception of the *Understand* level. The highest value in the *well written* parameter was achieved by the *Analyze* model with a value of 2.63. Considering the *understandability* of the fine-tuned models, the lowest also belongs to the *Understand* category. This is likely to happen due to the possible influence that the quality of written questions can have on *understandability*. The inverse scenario is not true, as the the highest *understandability* value is the *create* model (2.54) instead of the expected *apply* (2.52) model by a margin of 0.02. Still in Table 4.7, when considering *relevance*, the best questions belong to the *Evaluate* level, while the least relevant are from the *Remember* level. Looking at these results it is possible to understand that a well written question may influence its *understandability*, but this does not mean that a well written question (or an understandable question) is necessarily considered relevant. Regarding *answerability*, the highest scores are from the *Apply* and *Evaluate* models, with 60.42% and 58.33%, respectively. Contrasting this with the remaining parameters, it seems there is a relationship between *relevance* and *answerability* – relevant questions appear to be often answerable. A peculiar case is the *Create* model due to its low *answerability* value. This can be explained by the fact that the *Create* level is the highest of RBT's cognitive domain and it is therefore the most complicated level to generate. The nature of *Create* questions is often tied to innovation and going outside the scope of the provided

Table 4.7: Overall average scores obtained on all questions, for all parameters, detailed by taxonomy level of the Revised Bloom's Taxonomy. Scores for the first three parameters range between 1 and 3, where 3 is the best score possible. Answerability is a percentage of positive answers.

	Remember	Understand	Apply	Analyze	Evaluate	Create
Well Written	2.61 ±0.51	1.95 ±0.70	2.43 ±0.44	2.63 ±0.35	2.29 ±0.68	2.42 ±0.63
Understandability	2.42 ±0.70	2.13 ±0.56	2.29 ±0.64	2.52 ±0.58	2.16 ±0.64	2.54 ±0.60
Relevance	1.96 ±0.71	2.00 ±0.59	2.25 ±0.60	2.19 ±0.65	2.27 ±0.61	2.08 ±0.79
Answerability (%)	50.00	41.67	60.42	52.08	58.33	33.33

Table 4.8: Fleiss Kappa and Agreement percentages on all questions, for all parameters, detailed by taxonomy level of the Revised Bloom's Taxonomy.

Fleiss' Kappa	Remember	Understand	Apply	Analyze	Evaluate	Create
Well Written	0.42	0.15	0.40	0.63	0.27	0.11
Understandability	0.20	0.36	0.26	0.19	0.07	0.23
Relevance	0.21	0.33	0.33	0.24	0.18	- 0.09
Agreement Percentage	**Remember**	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Create**
Well Written	61.31	43.45	60.12	75.60	51.19	40.48
Understandability	46.43	57.14	50.60	45.83	38.10	48.81
Relevance	47.02	55.36	55.36	49.40	45.24	27.38

contexts by producing something new, reason why these questions are often well written and understandable, despite being as often unanswerable.

Looking at Fleiss' Kappa and the agreement percentage (shown in Table 4.8), the same pattern of Tables 4.4 and 4.6 happens – the *inter-rater agreement* is higher when assessing if a question is written properly. There are exceptions to this pattern in the *Understand* and *Create* models. A factor to be considered when analyzing the *inter-rater agreement* is how well the evaluators master English. From a scale of 1 to 5, the average English proficiency is placed at 3.75, but what truly matters is the standard deviation of 0.89. Since the evaluators do not possess similar proficiency levels, it is plausible that questions of higher complexity generate more doubt and disagreement in the grammatical interpretation of the question, hence reducing the value of Fleiss' Kappa. The highest value of kappa in the *well written* parameter is seen in the *Apply* model (with a value of 0.63 – substantial agreement [65]) while the lowest kappa value is shown in the *Create* model (with a value of 0.11 — slight agreement [65]).

Considering *understandability*, highest kappa value pertains to the *Understand* model (0.36) and the lowest kappa to the *Evaluate* model (0.07). The final outcome of this table comes with the *relevance* parameter, where the fine-tuned models with higher kappa values are *Understand* and *Apply*. Although the evaluators have a slight agreement for the *understand* model and a fair agreement in the *apply* model, they produced the same outcome in terms of agreement when it comes to *relevance*. Finally, the *relevance* of the *create* model is the only negative Fleiss' Kappa in all the studied scenarios. A negative kappa value translates to poor agreement (or no agreement whatsoever). This situation is to be expected when it comes to the *create* category. It is the product of the combination of the subjectivity of the generated questions (typically present in this level) with the subjectivity of the evaluated parameter itself along with the innovative and inventive purpose of a question of the *create* level.

5 Conclusions and Future Work

In this work, we tried to solve the problem of generating questions according to the guidelines provided by the Revised Bloom's Taxonomy (RBT). With this goal in mind, we attempt to mitigate the problem of having no adequate dataset by creating a dataset and using it to train a *question classifier* that can be applied to any question-based dataset. We use this classifier to label Question Answering in Context (QuAC) dataset with a suitable variety of questions, in order to use it as training data for the proposed *question generator*. The *question generator* is a Sequence-To-Sequence (Seq2Seq) model that follows and *encoder-decoder* framework with an *attention mechanism* and is first trained with Stanford Question Answering Dataset (SQuAD) in order to provide a base-model that is then *fine-tuned* with each of the dataset splits from QuAC, producing a model of each level of the RBT. The evaluation process of the proposed *question generator* suggests that it is indeed capable of producing questions of various levels of the taxonomy that are well written, substantially relevant and understandable, although there are limitating factors.

5.1 Question Classification

Taking into consideration the created classifier, it is possible to produce a working solution attending to the necessities of the RBT. After the experimental setup presented in Section 3, it was possible to confirm that a Support Vector Machine (SVM) approach achieves the highest value of accuracy of all the tested algorithms.

The proposed approach combines *LinearSVC* with a pipeline comprised of *preprocessing*, *feature extraction* (using *TF-IDF*) and *resampling* (using both *oversampling* and *undersampling*) strategies that further enhance its results. The final model was then *fine-tuned*, ultimately producing a Question Classification (QC) model that has a cross-validation accuracy of 80.74% and a test accuracy of 66.7%. An important benefit from this classifier is that it does not use any meticulously crafted heuristics and, as such, it requires little to no expertise to operate and apply to several domains. Although question classification was not not the main focus of this work, the developed *question classifier* was vital in the creation of a viable dataset that could support our solution to Question Generation (QG). In addition, it also important to evaluate the results obtained by our *question generator*, providing an additional perspective in a task difficult to evaluate.

5.2 Question Generation

The purpose of this work was to create valuable questions that abide by the RBT. The most difficult challenges were to produce relevant questions while not considering the answer as part of the question generation input, and also to work with the commonly used datasets in this task, due to the fact that they are typically tightly related to purely factual questions (which only encompass the two lower levels of the RBT). Despite the aforementioned challenges, the generator is able to output questions for all the levels of the taxonomy.

The results were evaluated automatically using BLEU-1, with the lowest value pertaining to the *Create* model (9.86) and the highest value being from the *Apply* model (15.53). Automatic evaluation was also directed at the output frequencies of each *fine-tuned* model (that is, how often a model produces questions of its taxonomy level). The least frequent model is *Evaluate*, with 67.5%, and the most successful model is *Create*, with 90%. Granted that there is an error associated with the classification itself, the remarks made in section **??** regarding the bias in the dataset for the *Create* level may also influence the obtained results for that model.

To further study the results produced by the generation, we performed a manual evaluation by surveying eight evaluators. Results show that most questions are well written (2.39 out of 3), understandable (2.33 out of 3), and relevant (2.13 out of 3). However, the generated questions only have an answer in the text 50.76% of the time, which is a possibly a consequence of leaving answers out of the generator's input as this process allows for generated questions to be more specific and text-related. In the context of this dissertation, the absence of the answer in the training/generation process may indeed result in questions that are vague but this also allows the creation of questions that are beyond the presented text, while still being relevant in the given context. Nonetheless, the generated questions are still relevant, meaning that they are worth finding the answer in order to better understand something in the text.

5.3 Limitations

Even though there are several positive results regarding the proposed approach, we believe that there are still some limitations that, when overcame, could greatly benefit the value of this contribution and even bring it to real-life scenarios. For instance, there are limitations associated with the vocabulary, as Seq2Seq models work with a limited vocabulary that does not adapt to new content. Strategies exist to mitigate this problem but, in the proposed solution, these were not applied, therefore, even though the proposed solution can be applied to different contexts, it does not provide guarantees of being generalizable to all contexts.

Another issue with the manual evaluation is that the calculated *inter-rater agreement* values are lower than desirable when it comes to the more subjective parameters such as *relevance* and *understandability*. This is probably due to the subjectivity associates with the task. Another limitation of this work is related to the dataset. Although the obtained results indicate that the proposed approach is capable of producing a variety of questions that respect the RBT, with more suitable datasets, the accomplishments attained in this research could improve significantly.

A final issue with the generation of questions concerns how to handle *anaphoras*. In the context of this dissertation, the created questionnaire had explicit rules that disregarded this issue by assuming that the presented personal pronouns were correct for any given situation.

5.4 Future Work

Considering the final state of this dissertation, and as pointed out in this section, there are some limitations that could be addressed. Regarding the vocabulary used by the Seq2Seq model (as well as the existance of *anaphoras*), we emphasize the need to implement a mechanism that can copy words from the context to the question [32, 34].

Another important line for future work is the improvement of the dataset. A way to address this research line consists in creating a larger dataset from several educational sources that is manually labeled according to the RBT. Still in the topic of datasets, it would be interesting if such a dataset also contemplated the *Knowledge dimension* of the RBT. To further elaborate, an additional difficulty with the current datasets is not in the provided contexts, but rather in the provided variety of questions.

A third suggestion would be to develop an evaluation strategy that could perform better in terms of interpreting the subjective parameters involved in appraising the quality of generated questions.

A final consideration for future work would be to adopt recent and prominent strategies such as the use of transformers-based models, like a Bidirectional Encoder Representations from Transformers (BERT) [66] pre-trained model that could be *fine-tuned* for the purpose of *question generation*.

References

[1] K. Mazidi and P. Tarau, “Automatic question generation: From nlu to nlg,” in *Proceedings of the 13th International Conference on Intelligent Tutoring Systems - Volume 9684*, ser. ITS 2016. Berlin, Heidelberg: Springer-Verlag, 2016, p. 23–33. [Online]. Available: https://doi.org/10.1007/978-3-319-39583-8_3

[2] S. K. Patil and M. M. Shreyas, “A Comparative Study of Question Bank Classification based on Revised Bloom’s Taxonomy using SVM and K-NN,” in *2017 2nd International Conference On Emerging Computation and Information Technologies, ICECIT 2017*, 2018, pp. 1–7.

[3] S. F. Kusuma, D. Siahaan, and U. L. Yuhana, “Automatic Indonesia’s questions classification based on bloom’s taxonomy using Natural Language Processing a preliminary study,” in *2015 International Conference on Information Technology Systems and Innovation (ICITSI)*, nov 2015, pp. 1–6.

[4] B. S. Bloom, M. D. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl, “The Classification of Educational Goals,” in *Taxonomy of educational objectives*. Longmans, Green, 1956, p. 207.

[5] D. R. Krathwohl, “A revision of bloom’s taxonomy: An overview,” *Theory into Practice*, vol. 41, no. 4, pp. 212–218, 2002.

[6] N. Yusof and C. J. Hui, “Determination of Bloom’s cognitive level of question items using artificial neural network,” in *2010 10th International Conference on Intelligent Systems Design and Applications*, nov 2010, pp. 866–870.

[7] C. Lee, T. Chen, L. Chen, P. Yang, and R. T. Tsai, “Automatic Question Generation from Children’s Stories for Companion Chatbot,” in *2018 IEEE International Conference on Information Reuse and Integration (IRI)*, jul 2018, pp. 491–494.

[8] R. Shah, D. Shah, and L. Kurup, “Automatic question generation for intelligent tutoring systems,” *2017 2nd International Conference on Communication Systems, Computing and IT Applications, CSCITA 2017 - Proceedings*, pp. 127–132, 2017.

[9] T. Dodiya and S. Jain, “Question classification for medical domain Question Answering system,” in *2016 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*, 2016, pp. 204–207.

[10] G. Näsström, “Interpretation of standards with Bloom’s revised taxonomy: A comparison of teachers and assessment experts,” *International Journal of Research and Method in Education*, vol. 32, no. 1, pp. 39–51, 2009.

[11] S. S. Haris and N. Omar, “Bloom’s taxonomy question categorization using rules and N-gram approach,” *Journal of Theoretical and Applied Information Technology*, 2015.

[12] A. Lajis, H. Md Nasir, and N. A. Aziz, “Proposed assessment framework based on bloom taxonomy cognitive competency: Introduction to programming,” in *ACM International Conference Proceeding Series*, 2018.

[13] N. Omar, S. S. Haris, R. Hassan, H. Arshad, M. Rahmat, N. F. A. Zainal, and R. Zulkifli, “Automated Analysis of Exam Questions According to Bloom’s Taxonomy,” *Procedia - Social and Behavioral Sciences*, vol. 59, pp. 297–303, 2012.

[14] M. Mohammed and N. Omar, “Question classification based on Bloom’s Taxonomy using enhanced TF-IDF,” *International Journal on Advanced Science, Engineering and Information Technology*, vol. 8, no. 4-2, pp. 1679–1685, 2018.

[15] M. Heilman and N. A. Smith, “Good question! Statistical ranking for question generation,” in *NAACL HLT 2010 - Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Proceedings of the Main Conference*, no. June, 2010, pp. 609–617.

[16] K. Dhole and C. D. Manning, “Syn-QG: Syntactic and shallow semantic rules for question generation,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, Jul. 2020, pp. 752–765. [Online]. Available: https://www.aclweb.org/anthology/2020.acl-main.69

[17] K. K. Schuler, “Verbnet: A broad-coverage, comprehensive verb lexicon,” Ph.D. dissertation, University of Pennsylvania, 2006. [Online]. Available: http://verbs.colorado.edu/ kipper/Papers/dissertation.pdf

[18] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, “Bleu: a method for automatic evaluation of machine translation,” in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics, Jul. 2002, pp. 311–318. [Online]. Available: https://www.aclweb.org/anthology/P02-1040

[19] P. Khullar, K. Rachna, M. Hase, and M. Shrivastava, “Automatic question generation using relative pronouns and adverbs,” in *Proceedings of ACL 2018, Student Research Workshop*. Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 153–158. [Online]. Available: https://www.aclweb.org/anthology/P18-3022

[20] R. Mitkov and L. A. Ha, “Computer-aided generation of multiple-choice tests,” in *Proceedings of the HLT-NAACL 03 Workshop on Building Educational Applications Using Natural Language Processing*, 2003, pp. 17–22.

[21] T. Raynaud, J. Subercaze, and F. Laforest, “Thematic Question Generation over Knowledge Bases,” *Proceedings - 2018 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2018*, pp. 1–8, 2019.

[22] A. Shirude, S. Totala, S. Nikhar, V. Attar, and J. Ramanand, “Automated Question Generation tool for structured data,” in *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, aug 2015, pp. 1546–1551.

[23] Q. Zhou, N. Yang, F. Wei, C. Tan, H. Bao, and M. Zhou, “Neural question generation from text: A preliminary study,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, X. Huang, J. Jiang, D. Zhao, Y. Feng, and Y. Hong, Eds., vol. 10619 LNAI. Cham: Springer International Publishing, 2018, pp. 662–671.

[24] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” in *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, ser. NIPS’14. Cambridge, MA, USA: MIT Press, 2014, p. 3104–3112.

[25] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, u. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Proceedings of the 31st International Con-*

ference on Neural Information Processing Systems, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 6000–6010.

[26] K. Cho, B. van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder–decoder approaches," in *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 103–111. [Online]. Available: https://www.aclweb.org/anthology/W14-4012

[27] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "SQuad: 100,000+ questions for machine comprehension of text," in *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings*, 2016.

[28] P. Rajpurkar, R. Jia, and P. Liang, "Know what you don't know: Unanswerable questions for SQuAD," in *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, 2018.

[29] J. Bao, Y. Gong, N. Duan, M. Zhou, and T. Zhao, "Question Generation With Doubly Adversarial Nets," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 11, pp. 2230–2239, nov 2018.

[30] A. Trischler, T. Wang, X. Yuan, J. Harris, A. Sordoni, P. Bachman, and K. Suleman, "Newsqa: A machine comprehension dataset," 2017.

[31] Z. Wang, A. S. Lan, W. Nie, A. E. Waters, P. J. Grimaldi, and R. G. Baraniuk, "QG-net: A Data-driven Question Generation Model for Educational Content," in *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*, ser. L@S '18. New York, NY, USA: ACM, 2018, pp. 7:1—-7:10. [Online]. Available: http://doi.acm.org/10.1145/3231644.3231654

[32] A. See, P. J. Liu, and C. D. Manning, "Get to the point: Summarization with pointer-generator networks," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vancouver, Canada: Association for Computational Linguistics, Jul. 2017, pp. 1073–1083. [Online]. Available: https://www.aclweb.org/anthology/P17-1099

[33] J. Pennington, R. Socher, and C. D. Manning, “Glove: Global vectors for word representation,” in *Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543. [Online]. Available: http://www.aclweb.org/anthology/D14-1162

[34] B. Liu, K. Lai, M. Zhao, Y. He, Y. Xu, D. Niu, and H. Wei, “Learning to generate questions by learning what not to generate,” *The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019*, pp. 1106–1118, 2019. [Online]. Available: http://dx.doi.org/10.1145/3308558.3313737

[35] J. Amidei, P. Piwek, and A. Willis, “Evaluation methodologies in automatic question generation 2013-2018,” in *Proceedings of the 11th International Conference on Natural Language Generation*. Tilburg University, The Netherlands: Association for Computational Linguistics, nov 2018, pp. 307–317. [Online]. Available: https://www.aclweb.org/anthology/W18-6537

[36] E. Choi, H. He, M. Iyyer, M. Yatskar, W.-t. Yih, Y. Choi, P. Liang, and L. Zettlemoyer, “QuAC: Question answering in context,” in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Brussels, Belgium: Association for Computational Linguistics, Oct.-Nov. 2018, pp. 2174–2184. [Online]. Available: https://www.aclweb.org/anthology/D18-1241

[37] V. A. Silva, I. I. Bittencourt, and J. C. Maldonado, “Automatic question classifiers: A systematic review,” *IEEE Transactions on Learning Technologies*, vol. 12, no. 4, pp. 485–502, 2019.

[38] X. Li and D. Roth, “Learning question classifiers: the role of semantic information,” *Nat. Lang. Eng.*, vol. 12, pp. 229–249, 2006.

[39] A. A. Yahya, A. Osman, A. Taleb, and A. A. Alattab, “Analyzing the Cognitive Level of Classroom Questions Using Machine Learning Techniques,” *Procedia - Social and Behavioral Sciences*, vol. 97, pp. 587–595, 2013.

[40] D. A. Abduljabbar and N. Omar, “Exam questions classification based on Bloom’s taxonomy cognitive level using classifiers combination,” *Journal of Theoretical and Applied Information Technology*, vol. 78, no. 3, pp. 447–455, 2015.

[41] M. Nikulin, “Chi-squared test for normality,” in *Proceedings of the International Vilnius Conference on Probability Theory and Mathematical Statistics*, vol. 2, no. 1, 1973, pp. 119–122.

[42] A. Osadi, N. Fernando, and V. Welgama, “Ensemble Classifier based Approach for Classification of Examination Questions into Bloom’s Taxonomy Cognitive Levels,” *International Journal of Computer Applications*, vol. 162, pp. 975–8887, 2017.

[43] A. Osman and A. Yahya, “Classifications of Exam Questions Using Linguistically-Motivated Features: A Case Study Based on Bloom’s Taxonomy,” in *The Sixth International Arab Conference on Quality Assurance in Higher Education (IACQA’ 2016)*, 2016.

[44] M. Honnibal and I. Montani, “spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing,” 2017.

[45] A. Fernández, S. García, F. Herrera, and N. V. Chawla, “Smote for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary,” *J. Artif. Int. Res.*, vol. 61, no. 1, p. 863–905, Jan. 2018.

[46] C.-C. Chang and C.-J. Lin, “Libsvm: A library for support vector machines,” *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, May 2011. [Online]. Available: https://doi.org/10.1145/1961189.1961199

[47] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, “Liblinear: A library for large linear classification,” *J. Mach. Learn. Res.*, vol. 9, p. 1871–1874, Jun. 2008.

[48] H. Zhang, “The optimality of naïve bayes,” in *In FLAIRS2004 conference*, 2004.

[49] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, “Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization,” *ACM Trans. Math. Softw.*, vol. 23, no. 4, p. 550–560, Dec. 1997. [Online]. Available: https://doi.org/10.1145/279232.279236

[50] T. Cover and P. Hart, “Nearest neighbor pattern classification,” *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[51] J. H. Friedman, “Stochastic gradient boosting,” *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367 – 378, 2002, nonlinear Methods and Data Mining. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167947301000652

[52] J. Friedman, “Greedy function approximation: A gradient boosting machine,” *The Annals of Statistics*, vol. 29, 11 2000.

[53] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119 – 139, 1997. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S002200009791504X

[54] J. D. M. Rennie, L. Shih, J. Teevan, and D. R. Karger, “Tackling the poor assumptions of naive bayes text classifiers,” in *Proceedings of the Twentieth International Conference on International Conference on Machine Learning*, ser. ICML’03. AAAI Press, 2003, p. 616–623.

[55] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.

[56] K. N. Stevens, P. M. Eden, I. Pollack, L. Ficks, I. O. Multidimensional, A. Displays, C. W. Ericsen, M. S. Differences, and P. E. Hart, “Nearest neighbor pattern classification,” *IEEE Transactions on Information Theory*, 1953.

[57] H. Zou, J. Zhu, S. Rosset, and T. Hastie, “Multi-class adaboost,” *Statistics and its Interface*, vol. 2, pp. 349–360, 2009.

[58] M. C. Burkhart and K. Shan, “Deep low-density separation for semi-supervised classification,” in *Computational Science – ICCS 2020*, V. V. Krzhizhanovskaya, G. Závodszky, M. H. Lees, J. J. Dongarra, P. M. A. Sloot, S. Brissos, and J. Teixeira, Eds. Cham: Springer International Publishing, 2020, pp. 297–311.

[59] I. Aldabe, M. L. de Lacalle, M. Maritxalar, E. Martinez, and L. Uria, “Arikiturri: An automatic question generator based on corpora and nlp techniques,” in *Intelligent Tutoring Systems*, M. Ikeda, K. D. Ashley, and T.-W. Chan, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 584–594.

[60] M. Schuster and K. Paliwal, “Bidirectional recurrent neural networks,” *Trans. Sig. Proc.*, vol. 45, no. 11, p. 2673–2681, Nov. 1997. [Online]. Available: https://doi.org/10.1109/78.650093

[61] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, p. 1735–1780, Nov. 1997. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735

[62] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” 2016.

[63] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[64] J. Fleiss *et al.*, “Measuring nominal scale agreement among many raters,” *Psychological Bulletin*, vol. 76, no. 5, pp. 378–382, 1971.

[65] J. R. Landis and G. G. Koch, “The measurement of observer agreement for categorical data,” *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977. [Online]. Available: http://www.jstor.org/stable/2529310

[66] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” 2019.

A Questionnaire

Inquérito

O meu nome é Gonçalo Correia, estudante de mestrado em Engenharia Informática no Iscte - Instituto Universitário de Lisboa, orientada pelo professor Ricardo Ribeiro e pelo mestre Hugo Rodrigues.

Este inquérito surge no âmbito da minha dissertação de mestrado - "A Neural Network Approach for Automatic Question Generation using Bloom's Taxonomy" com o objetivo de apurar a qualidade da solução proposta.

Neste documento encontram-se, em inglês, 37 parágrafos. Cada parágrafo é seguido por uma questão a avaliar em quatro aspetos:

Bem escrita– Se apresenta erros ortográficos (1 – Erros que tornam a pergunta difícil de compreender; 2 – Gaffes e/ou palavras repetidas; 3 – Sem erros).

Compreensibilidade - Se a pergunta é percetivel e é possível compreender o seu propósito (1 - incompreensível, 2 - é compreensível mas ambígua, 3 - é perfeitamente compreensível)

Relevância – Se a pergunta está relacionada com o texto (1 - não relacionada; 2 - há relação, mas não evidente; 3 - relação evidente)

A resposta está no texto - Se a pergunta tem resposta possível com base no parágrafo (Sim/Não)

É expectável que o inquérito demore cerca de 45 minutos a responder.

NOTA: Deverá considerar as questões com pronome (He, She, It..) como uma referência correta, a menos que a pessoa a quem o pronome se refira não seja do género referido pelo pronome ou não exista.

*Obrigatório

1. Idade *

____

2. Nível de ensino *

Marcar apenas uma oval.

○ Básico
○ Secundário
○ Ensino Superior

3. Área de formação *

Marcar apenas uma oval.

○ Ciências
○ Humanidades
○ Artes
○ Outra

4. Nível de Proficiência na Língua Inglesa *

Marcar apenas uma oval.

	1	2	3	4	5	
Baixo	○	○	○	○	○	Superior

Questão 1

"In August 2003, Lopez starred opposite Affleck in the romantic comedy Gigli. The film was a box office bomb and is considered one of the worst films of all time. The film's poor reception was attributed to negative press preceding its release, as well as the media attention surrounding Lopez and Affleck's engagement which largely overshadowed the film. Lopez would later describe this as the lowest point of her career, saying "[It] was very tough", "the tabloid press had just come into existence at the time, so I was like a poster child for that moment." In October of that year, she released her next fragrance, Still Jennifer Lopez. Lopez also launched her next fashion label, Sweetface. It was described by Andy Hilfiger as a ""more intellectual, more inspirational collection than J-Lo by Jennifer Lopez. Less sporty, more suede."" Lopez's clothing lines and two fragrances generated over $300 million in revenue throughout 2004, which made her the 19th richest person under 40. In March 2004, Lopez had a minor role in the film Jersey Girl, alongside Affleck. Her character, Gertrude Steiney, dies during childbirth within the first 15 minutes of the film. From the intense media scrutiny following the couple's break-up, it was noted that "they may need to put Lopez in a coffin on the poster if they want anyone to come"".

[QUESTION]: What was her role in the movie?

5. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

6. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

7. Relevância *

Marcar apenas uma oval.

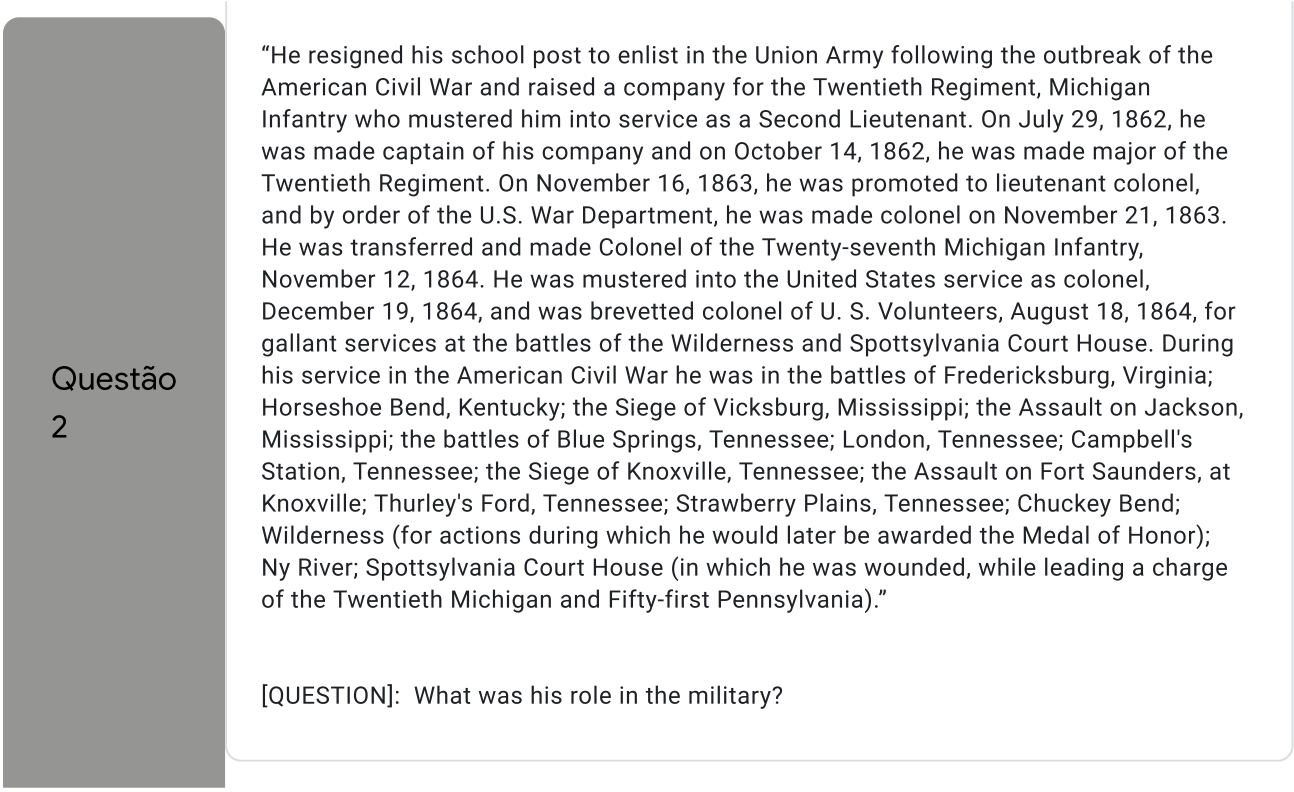
	1	2	3	
Não relacionada	○	○	○	Relação evidente

8. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não



Questão 2

“He resigned his school post to enlist in the Union Army following the outbreak of the American Civil War and raised a company for the Twentieth Regiment, Michigan Infantry who mustered him into service as a Second Lieutenant. On July 29, 1862, he was made captain of his company and on October 14, 1862, he was made major of the Twentieth Regiment. On November 16, 1863, he was promoted to lieutenant colonel, and by order of the U.S. War Department, he was made colonel on November 21, 1863. He was transferred and made Colonel of the Twenty-seventh Michigan Infantry, November 12, 1864. He was mustered into the United States service as colonel, December 19, 1864, and was brevetted colonel of U. S. Volunteers, August 18, 1864, for gallant services at the battles of the Wilderness and Spottsylvania Court House. During his service in the American Civil War he was in the battles of Fredericksburg, Virginia; Horseshoe Bend, Kentucky; the Siege of Vicksburg, Mississippi; the Assault on Jackson, Mississippi; the battles of Blue Springs, Tennessee; London, Tennessee; Campbell's Station, Tennessee; the Siege of Knoxville, Tennessee; the Assault on Fort Saunders, at Knoxville; Thurley's Ford, Tennessee; Strawberry Plains, Tennessee; Chuckey Bend; Wilderness (for actions during which he would later be awarded the Medal of Honor); Ny River; Spottsylvania Court House (in which he was wounded, while leading a charge of the Twentieth Michigan and Fifty-first Pennsylvania).”

[QUESTION]: What was his role in the military?

9. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

10. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

11. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

12. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não

Questão 3

"His first book, La Mediterranee et le Monde Mediterraneen a l'Epoque de Philippe II (1949) (The Mediterranean and the Mediterranean World in the Age of Philip II) was his most influential. For Braudel there is no single Mediterranean Sea. There are many seas--indeed a ""vast, complex expanse"" within which men operate. Life is conducted on the Mediterranean: people travel, fish, fight wars, and drown in its various contexts. And the sea articulates with the plains and islands. Life on the plains is diverse and complex; the poorer south is affected by religious diversity (Catholicism and Islam), as well as by intrusions - both cultural and economic - from the wealthier north. In other words, the Mediterranean cannot be understood independently from what is exterior to it. Any rigid adherence to boundaries falsifies the situation. The first level of time, geographical time, is that of the environment, with its slow, almost imperceptible change, its repetition and cycles. Such change may be slow, but it is irresistible. The second level of time comprises long-term social, economic, and cultural history, where Braudel discusses the Mediterranean economy, social groupings, empires and civilizations. Change at this level is much more rapid than that of the environment; Braudel looks at two or three centuries in order to spot a particular pattern, such as the rise and fall of various aristocracies. The third level of time is that of events (histoire evenementielle). This is the history of individuals with names. This, for Braudel, is the time of surfaces and deceptive effects. It is the time of the ""courte duree"" proper and it is the focus of Part 3 of The Mediterranean which treats of ""events, politics and people."" Braudel's Mediterranean is centered on the sea, but just as important, it is also the desert and the mountains. The desert creates a nomadic form of social organization where the whole community moves; mountain life is sedentary. Transhumance -- that is, the movement from the mountain to the plain, or vice versa in a given season -- is also a persistent part of Mediterranean existence. Braudel's vast, panoramic view used insights from other social sciences, employed the concept of the longue duree, and downplayed the importance of specific events. It was widely admired, but most historians did not try to replicate it and instead focused on their specialized monographs. The book firmly launched the study of the Mediterranean and dramatically raised the worldwide profile of the Annales School."

[QUESTION]: What was his argument?

13. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

14. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

15. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

16. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não

Questão 4

“Branson started his record business from the church where he ran Student magazine. He interviewed several prominent personalities of the late 1960s for the magazine including Mick Jagger and R. D. Laing. Branson advertised popular records in Student, and it was an overnight success. Trading under the name ""Virgin"", he sold records for considerably less than the ""High Street"" outlets, especially the chain W. H. Smith. Branson once said, ""There is no point in starting your own business unless you do it out of a sense of frustration."" The name ""Virgin"" was suggested by one of Branson's early employees because they were all new at business. At the time, many products were sold under restrictive marketing agreements that limited discounting, despite efforts in the 1950s and 1960s to limit so-called resale price maintenance. Branson eventually started a record shop in Oxford Street in London. In 1971, he was questioned in connection with the selling of records in Virgin stores that had been declared export stock. The matter was never brought before a court because Branson agreed to repay any unpaid VAT of 33% and a PS70,000 fine. His parents re-mortgaged the family home in order to help pay the settlement. Earning enough money from his record store, Branson in 1972 launched the record label Virgin Records with Nik Powell, and bought a country estate north of Oxford in which he installed a residential recording studio, The Manor Studio. He leased studio time to fledgling artists, including multi-instrumentalist Mike Oldfield, whose debut album Tubular Bells (1973) was the first release for Virgin Records and became a chart-topping best-seller. Virgin signed such controversial bands as the Sex Pistols, which other companies were reluctant to sign. Virgin Records would go on to sign other artists including the Rolling Stones, Peter Gabriel, UB40, Steve Winwood and Paula Abdul, and to become the world's largest independent record label. It also won praise for exposing the public to such obscure avant-garde music as Faust and Can. Virgin Records also introduced Culture Club to the music world. In 1982, Virgin purchased the gay nightclub Heaven. In 1991, in a consortium with David Frost, Branson made an unsuccessful bid for three ITV franchisees under the CPV-TV name. The early 1980s also saw his only attempt as a producer--on the novelty record ""Baa, Baa, Black Sheep"", by Singing Sheep in association with Doug McLean and Grace McDonald. The recording was a series of sheep baa-ing along to a drum-machine-produced track and reached number 42 in the UK charts in 1982. In 1992, to keep his airline company afloat, Branson sold the Virgin label to EMI for PS500 million. Branson said that he wept when the sale was completed because the record business had been the very start of the Virgin empire. He created V2 Records in 1996 in order to re-enter the music business, owning 5% himself. “

[QUESTION]: How did he do in the 1980’s?

17. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

18. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

19. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

20. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não

Questão 5

"In 1963, Major League Baseball expanded the strike zone. Compared to the previous season, National League walks fell 13 percent, strikeouts increased six percent, the league batting average fell from .261 to .245, and runs fell 15 percent. Koufax, who had reduced his walks allowed per nine innings to 3.4 in 1961 and 2.8 in 1962, reduced his walk rate further to 1.7 in 1963, which ranked fifth in the league. The top pitchers of the era - Don Drysdale, Juan Marichal, Jim Bunning, Bob Gibson, Warren Spahn, and above all Koufax - significantly reduced the walks-given-up-to-batters-faced ratio for 1963, and subsequent years. On May 11, Koufax no-hit the San Francisco Giants 8-0, besting future Hall of Fame pitcher Juan Marichal--himself a no-hit pitcher a month later, on June 15. Koufax carried a perfect game into the eighth inning against the powerful Giants lineup, including future Hall of Famers Willie Mays, Willie McCovey, and Orlando Cepeda. He walked Ed Bailey on a 3-and-2 pitch in the 8th, and pinch-hitter McCovey on four pitches in the 9th, before closing out the game. As the Dodgers won the pennant, Koufax won the pitchers' Triple Crown, leading the league in wins (25), strikeouts (306) and ERA (1.88). Koufax threw 11 shutouts, setting a new post-1900 record for shutouts by a left-handed pitcher that stands to this day (the previous record of 10 shutouts had been held by Carl Hubbell for 30 years). Only Bob Gibson, a right-hander, has thrown more shutouts (13) since, and that was in 1968, ""the year of the pitcher."" Koufax won the NL MVP Award and the Hickok Belt, and was the first-ever unanimous selection for the Cy Young Award. Facing the Yankees in the 1963 World Series, Koufax beat Whitey Ford 5-2 in Game 1 and struck out 15 batters -- including the first 5, breaking Carl Erskine's decade-old record of 14 (Gibson would break Koufax's record by striking out 17 Detroit Tigers in the 1968 World Series opener). After seeing Koufax's Game 1 performance, Yogi Berra said, ""I can see how he won 25 games. What I don't understand is how he lost five,"" to which Maury Wills responded, ""He didn't. We lost them for him."" In Game 4, Koufax completed the Dodgers' series sweep with a 2-1 victory over Ford, clinching the Series MVP Award for his performance."

[QUESTION]: Did they do in the playoffs?

21. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

22. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

23. Relevância *

Marcar apenas uma oval.

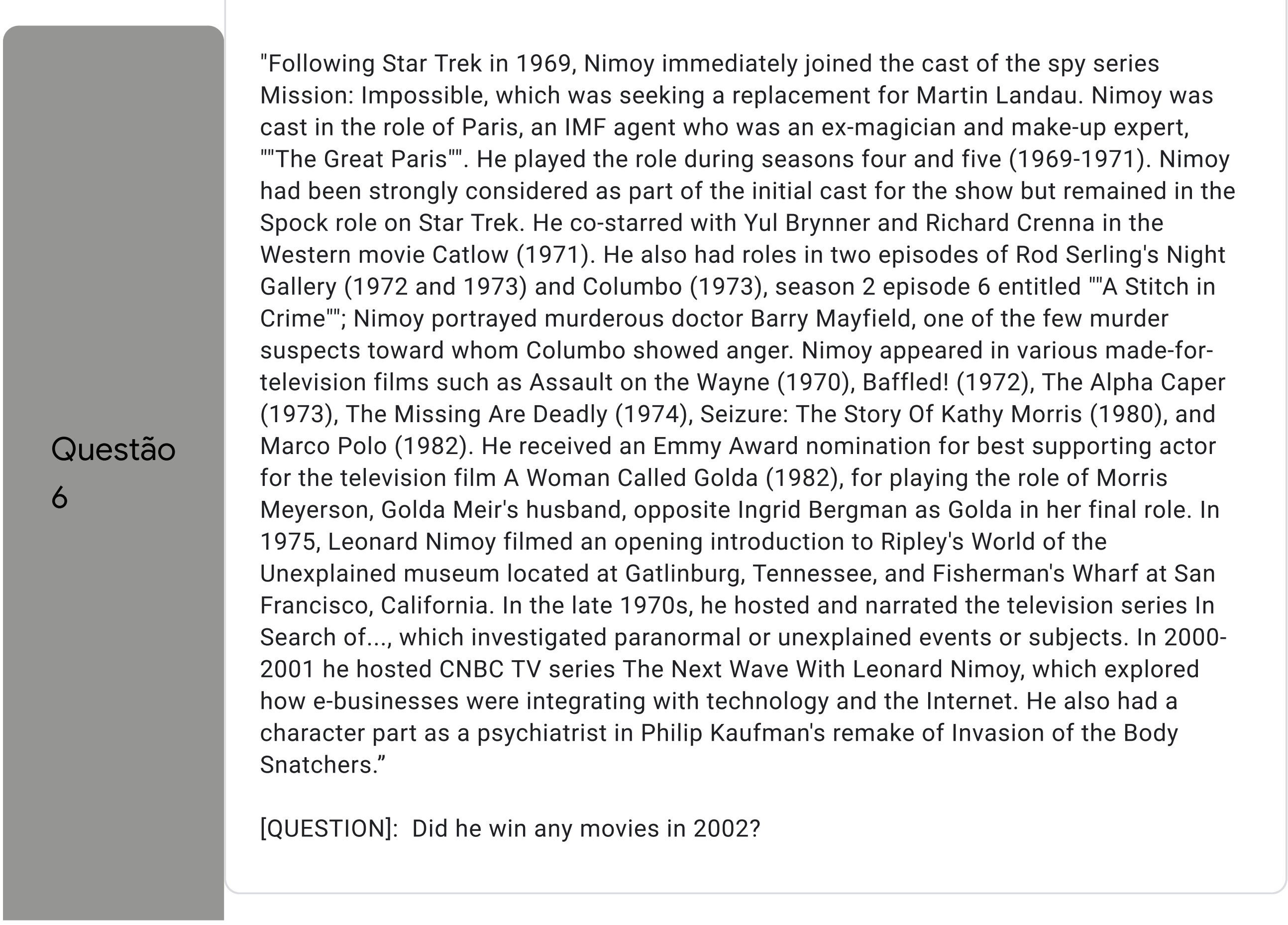
	1	2	3	
Não relacionada	○	○	○	Relação evidente

24. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não



Questão 6

"Following Star Trek in 1969, Nimoy immediately joined the cast of the spy series Mission: Impossible, which was seeking a replacement for Martin Landau. Nimoy was cast in the role of Paris, an IMF agent who was an ex-magician and make-up expert, ""The Great Paris"". He played the role during seasons four and five (1969-1971). Nimoy had been strongly considered as part of the initial cast for the show but remained in the Spock role on Star Trek. He co-starred with Yul Brynner and Richard Crenna in the Western movie Catlow (1971). He also had roles in two episodes of Rod Serling's Night Gallery (1972 and 1973) and Columbo (1973), season 2 episode 6 entitled ""A Stitch in Crime""; Nimoy portrayed murderous doctor Barry Mayfield, one of the few murder suspects toward whom Columbo showed anger. Nimoy appeared in various made-for-television films such as Assault on the Wayne (1970), Baffled! (1972), The Alpha Caper (1973), The Missing Are Deadly (1974), Seizure: The Story Of Kathy Morris (1980), and Marco Polo (1982). He received an Emmy Award nomination for best supporting actor for the television film A Woman Called Golda (1982), for playing the role of Morris Meyerson, Golda Meir's husband, opposite Ingrid Bergman as Golda in her final role. In 1975, Leonard Nimoy filmed an opening introduction to Ripley's World of the Unexplained museum located at Gatlinburg, Tennessee, and Fisherman's Wharf at San Francisco, California. In the late 1970s, he hosted and narrated the television series In Search of..., which investigated paranormal or unexplained events or subjects. In 2000-2001 he hosted CNBC TV series The Next Wave With Leonard Nimoy, which explored how e-businesses were integrating with technology and the Internet. He also had a character part as a psychiatrist in Philip Kaufman's remake of Invasion of the Body Snatchers."

[QUESTION]: Did he win any movies in 2002?

25. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

26. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

27. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

28. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não

Questão 7

“Hartson turned professional in 1992 and made his Luton Town first team debut as a substitute in a 1-0 League Cup defeat to Cambridge United on 24 August 1993. In January 1995, at 19, he joined Arsenal for PS2.5 million, a British record fee for a teenage player at the time. Along with Chris Kiwomya, Hartson was one of George Graham's last signings before the manager's sacking in February 1995. He made his Arsenal debut on 14 January 1995, a 1-1 home draw with Everton, and scored his first goal for the club the following week, the only goal in a 1-0 away win at Coventry City. He was a regular for the remainder of his first season, a highlight of which was scoring Arsenal's 75th-minute equaliser in the 1995 UEFA Cup Winners' Cup Final against Real Zaragoza; however a last-minute goal from 40 yards by Nayim over David Seaman meant Arsenal lost the game 2-1. He was strike-partner to Ian Wright, being favoured ahead of Kevin Campbell to fill the gap left by the injured Alan Smith, who would retire at the end of the season. Following the signing of Dennis Bergkamp, who was preferred up front to partner Wright, Hartson went on to feature under Graham's successors Bruce Rioch and Arsene Wenger. With Wenger wanting him to stay at the club, Hartson though in February 1997 linked up with West Ham United in a PS3.2 million deal. At the time, he was the most expensive player to be signed by West Ham.The deal was initially reported to be worth PS5 million. In total, Hartson played 53 times for Arsenal, scoring 14 goals.”

[QUESTION]: How many goals did he make?

29. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

30. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

31. Relevância *

Marcar apenas uma oval.

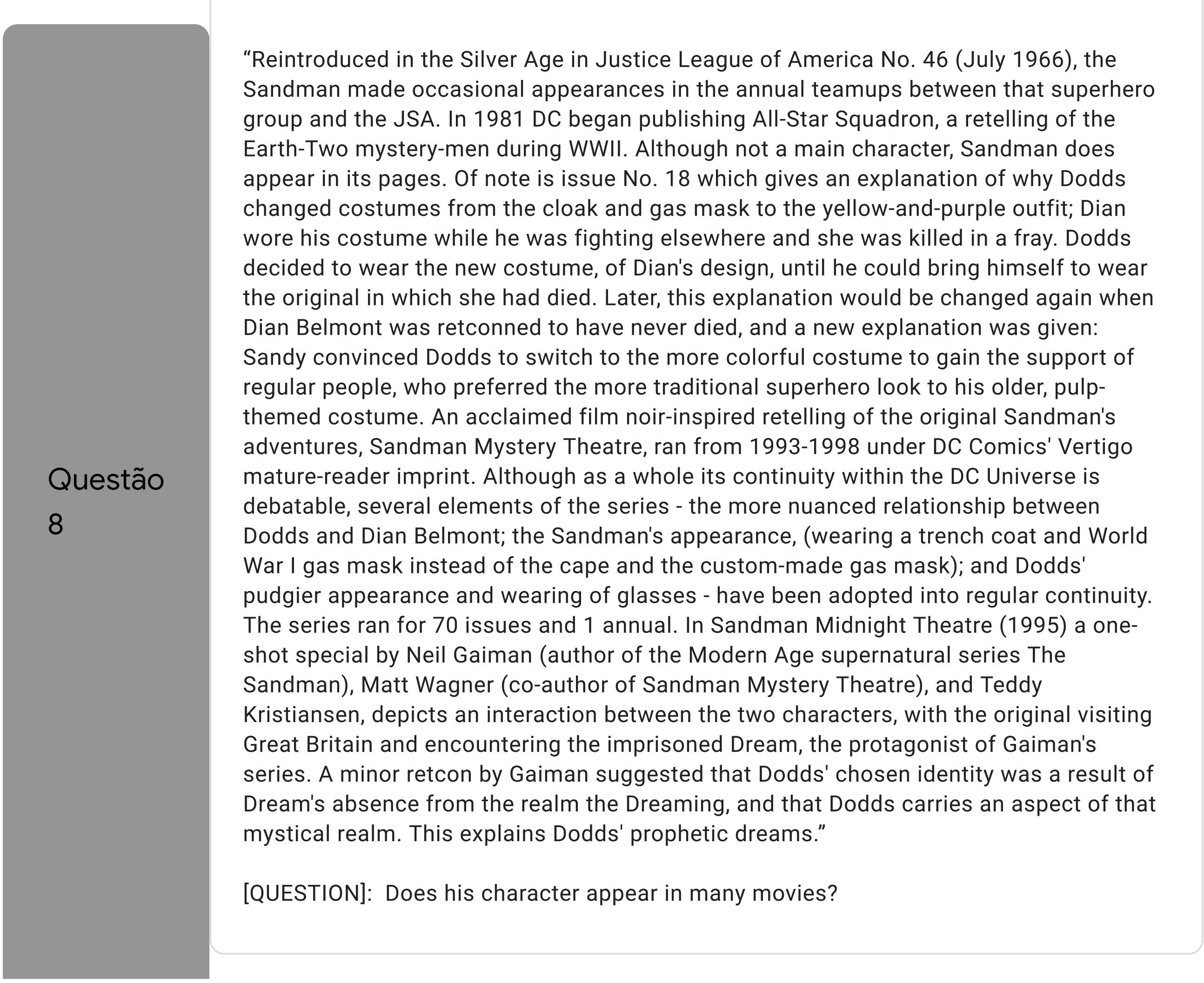
	1	2	3	
Não relacionada	○	○	○	Relação evidente

32. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não



Questão 8

“Reintroduced in the Silver Age in Justice League of America No. 46 (July 1966), the Sandman made occasional appearances in the annual teamups between that superhero group and the JSA. In 1981 DC began publishing All-Star Squadron, a retelling of the Earth-Two mystery-men during WWII. Although not a main character, Sandman does appear in its pages. Of note is issue No. 18 which gives an explanation of why Dodds changed costumes from the cloak and gas mask to the yellow-and-purple outfit; Dian wore his costume while he was fighting elsewhere and she was killed in a fray. Dodds decided to wear the new costume, of Dian's design, until he could bring himself to wear the original in which she had died. Later, this explanation would be changed again when Dian Belmont was retconned to have never died, and a new explanation was given: Sandy convinced Dodds to switch to the more colorful costume to gain the support of regular people, who preferred the more traditional superhero look to his older, pulp-themed costume. An acclaimed film noir-inspired retelling of the original Sandman's adventures, Sandman Mystery Theatre, ran from 1993-1998 under DC Comics' Vertigo mature-reader imprint. Although as a whole its continuity within the DC Universe is debatable, several elements of the series - the more nuanced relationship between Dodds and Dian Belmont; the Sandman's appearance, (wearing a trench coat and World War I gas mask instead of the cape and the custom-made gas mask); and Dodds' pudgier appearance and wearing of glasses - have been adopted into regular continuity. The series ran for 70 issues and 1 annual. In Sandman Midnight Theatre (1995) a one-shot special by Neil Gaiman (author of the Modern Age supernatural series The Sandman), Matt Wagner (co-author of Sandman Mystery Theatre), and Teddy Kristiansen, depicts an interaction between the two characters, with the original visiting Great Britain and encountering the imprisoned Dream, the protagonist of Gaiman's series. A minor retcon by Gaiman suggested that Dodds' chosen identity was a result of Dream's absence from the realm the Dreaming, and that Dodds carries an aspect of that mystical realm. This explains Dodds' prophetic dreams.”

[QUESTION]: Does his character appear in many movies?

33. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

34. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

35. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

36. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim
○ Não

Questão 9

“Both he and his next opponent, Tyrone Booze, moved up to the cruiserweight division for their fight on July 20, 1985, in Norfolk, Virginia. Holyfield won an eight-round decision over Booze. Evander went on to knock out Rick Myers in the first round on August 29 in Holyfield's hometown of Atlanta. On October 30 in Atlantic City he knocked out opponent Jeff Meachem in five rounds, and his last fight for 1985 was against Anthony Davis on December 21 in Virginia Beach, Virginia. He won by knocking out Davis in the fourth round. He began 1986 with a knockout in three rounds over former world cruiserweight challenger Chisanda Mutti, and proceeded to beat Jessy Shelby and Terry Mims before being given a world title try by the WBA Cruiserweight Champion Dwight Muhammad Qawi. In what was called by The Ring as the best cruiserweight bout of the 1980s, Holyfield became world champion by defeating Qawi by a narrow 15 round split decision. He culminated 1986 with a trip to Paris, France, where he beat Mike Brothers by a knockout in three, in a non-title bout. In 1987, he defended his title against former Olympic teammate and Gold medal winner Henry Tillman, who had beaten Mike Tyson twice as an amateur. He retained his belt, winning by seventh-round knockout, and then went on to unify his WBA belt with the IBF belt held by Ricky Parkey, knocking Parkey out in three rounds. For his next bout, he returned to France, where he retained the title with an eleven-round knockout against former world champion Ossie Ocasio. In his last fight of 1987, he offered Muhammad Qawi a rematch and, this time, he beat Qawi by a knockout in only four rounds. 1988 was another productive year for Holyfield; he started by becoming the first universally recognized World Cruiserweight Champion after defeating the Lineal & WBC Champion Carlos De Leon at Las Vegas. The fight was stopped after eight rounds.”

[QUESTION]: How many rounds did they fight?

37. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

38. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

39. Relevância *

Marcar apenas uma oval.

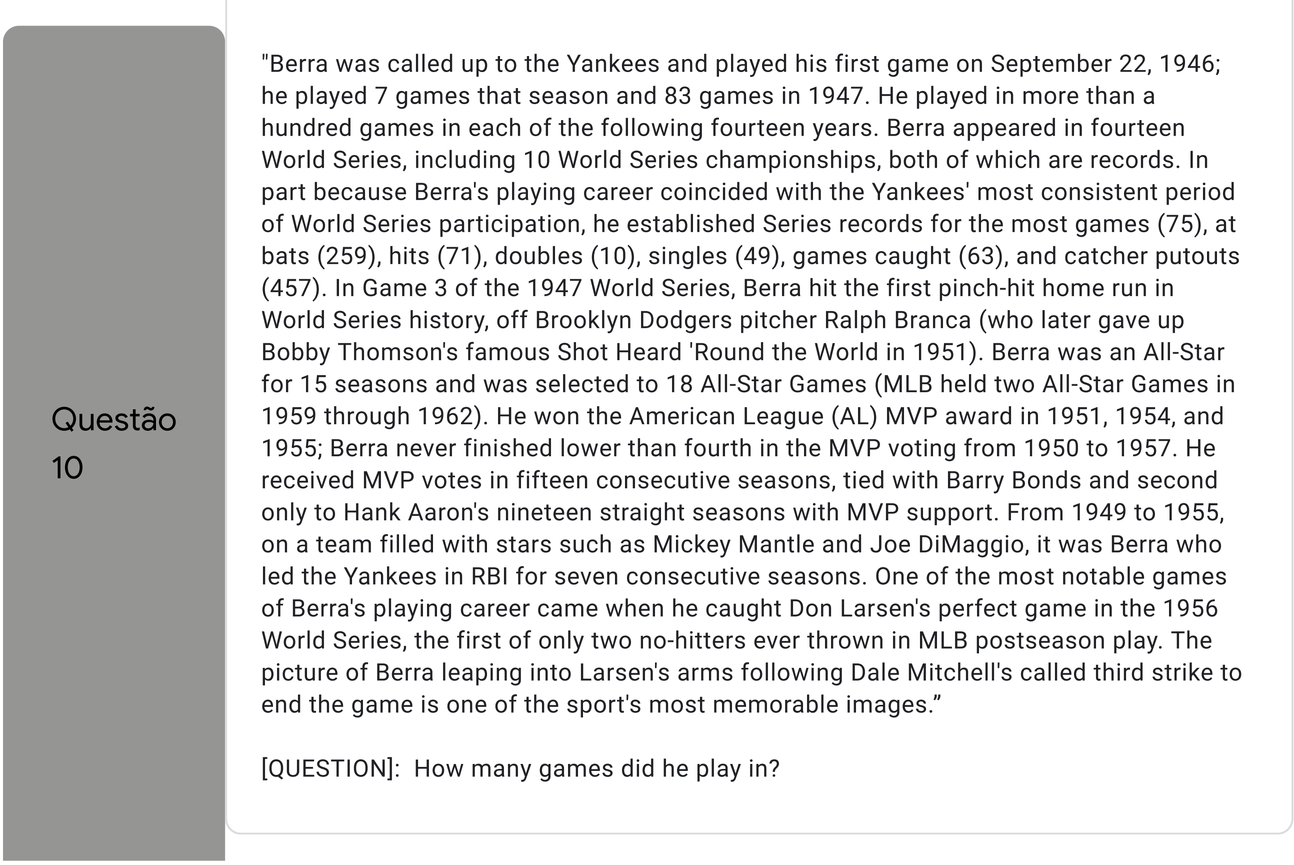
	1	2	3	
Não relacionada	○	○	○	Relação evidente

40. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não



Questão 10

"Berra was called up to the Yankees and played his first game on September 22, 1946; he played 7 games that season and 83 games in 1947. He played in more than a hundred games in each of the following fourteen years. Berra appeared in fourteen World Series, including 10 World Series championships, both of which are records. In part because Berra's playing career coincided with the Yankees' most consistent period of World Series participation, he established Series records for the most games (75), at bats (259), hits (71), doubles (10), singles (49), games caught (63), and catcher putouts (457). In Game 3 of the 1947 World Series, Berra hit the first pinch-hit home run in World Series history, off Brooklyn Dodgers pitcher Ralph Branca (who later gave up Bobby Thomson's famous Shot Heard 'Round the World in 1951). Berra was an All-Star for 15 seasons and was selected to 18 All-Star Games (MLB held two All-Star Games in 1959 through 1962). He won the American League (AL) MVP award in 1951, 1954, and 1955; Berra never finished lower than fourth in the MVP voting from 1950 to 1957. He received MVP votes in fifteen consecutive seasons, tied with Barry Bonds and second only to Hank Aaron's nineteen straight seasons with MVP support. From 1949 to 1955, on a team filled with stars such as Mickey Mantle and Joe DiMaggio, it was Berra who led the Yankees in RBI for seven consecutive seasons. One of the most notable games of Berra's playing career came when he caught Don Larsen's perfect game in the 1956 World Series, the first of only two no-hitters ever thrown in MLB postseason play. The picture of Berra leaping into Larsen's arms following Dale Mitchell's called third strike to end the game is one of the sport's most memorable images."

[QUESTION]: How many games did he play in?

41. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

42. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

43. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

44. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não

Questão 11

"After the success of Parachutes, Coldplay returned to the studio in September 2001 to begin work on their second album, A Rush of Blood to the Head, once again with Ken Nelson producing. Since the band had never stayed in London before, they had trouble focusing. They decided to relocate in Liverpool, where they recorded some of the songs on Parachutes. Once there, vocalist Chris Martin said that they became obsessed with recording. ""In My Place"" was the first song recorded for the album. The band released it as the album's lead single because it was the track that made them want to record a second album, following a ""strange period of not really knowing what we were doing"" three months after the success of Parachutes. According to Martin ""one thing kept us going: recording 'In My Place'. Then other songs started coming."" The band wrote more than 20 songs for the album. Some of their new material, including ""In My Place"" and ""Animals"", was played live while the band was still touring Parachutes. The album's title was revealed through a post on the band's official website. The album was released in August 2002 and spawned several popular singles, including ""In My Place"", ""Clocks"", and the ballad ""The Scientist"". The latter was inspired by George Harrison's ""All Things Must Pass"", which was released in 1970. Coldplay toured from June 2002 to September 2003 for the A Rush of Blood to the Head Tour. They visited five continents, including co-headlining festival dates at Glastonbury Festival, V2003 and Rock Werchter. Many concerts showcased elaborate lighting and individualised screens reminiscent of U2's Elevation Tour and Nine Inch Nails' Fragility Tour. During the extended tour, Coldplay recorded a live DVD and CD, Live 2003, at Sydney's Hordern Pavilion. At the 2003 Brit Awards held at Earls Court, London, Coldplay received awards for Best British Group, and Best British Album. On 28 August 2003, Coldplay performed ""The Scientist"" at the 2003 MTV Video Music Awards at the Radio City Music Hall in New York City, and won three awards. In December 2003, readers of Rolling Stone chose Coldplay as the best artist and the best band of the year. At that time the band covered The Pretenders' 1983 song ""2000 Miles"" (which was made available for download on their official website). ""2000 Miles"" was the top selling UK download that year, with proceeds from the sales donated to Future Forests and Stop Handgun Violence campaigns. A Rush of Blood to the Head won the Grammy Award for Best Alternative Music Album at the 2003 Grammy Awards. At the 2004 Grammy Awards, Coldplay earned Record of the Year for ""Clocks""."

[QUESTION]: When was the album released?

45. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

46. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

47. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

48. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não

Questão 12

“On November 15, 2016, the band announced that in celebration of its upcoming fifteenth year anniversary, they would be embarking on the Quince Anos Tour in March and April 2017, with support from Counterparts, Movements, and Like Pacific. To commemorate the event, the band performed it's 2006 album Still Searching in full, alongside a collection of career spanning songs. On the same day, the band announced the release of their long teased acoustic EP, ""In Your Absence"". It features 3 brand new songs, alongside acoustic renditions of ""Lost and Found"" from Still Searching, and ""Family Tradition"" from Life Is Not A Waiting Room. A music video was released for the lead single, ""Jets to Peru"", on January 26, 2017. The EP released on March 3, 2017 alongside the beginning of the Quince Anos Tour. The band entered the studio with Saosin guitarist Beau Burchell, who also handled recording duties on In Your Absence, to begin recording their seventh full-length album on June 27, 2017, to be titled If There Is Light, It Will Find You. Nielsen commented that the album would feature a style more akin to earlier releases, such as Let It Enfold You. The album will be written entirely by Nielsen. On August 2, it was revealed that former drummer Dan Trapp would be performing drums on the album, although current drummer Chris Hornbrook would still be performing and touring with the band. However on January 8, 2018, Hornbrook announced his departure from the band. Hornbrook had been touring with Dhani Harrison during the recording sessions. On February 1, 2018, Steve Carey of The Color Morale was announced as the bands new drummer, following the announcement that The Color Morale would be entering a hiatus. On November 30, the lead single ""Double Cross"" was released. A second single, ""Gold Jacket, Green Jacket..."" was released on January 11, 2018. On February 1, 2018, a third single, ""New Jersey Makes, the World Takes"" was released. The album was released on February 16th, 2018. "

[QUESTION]: Did they have any hits?

49. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

50. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

51. Relevância *

Marcar apenas uma oval.

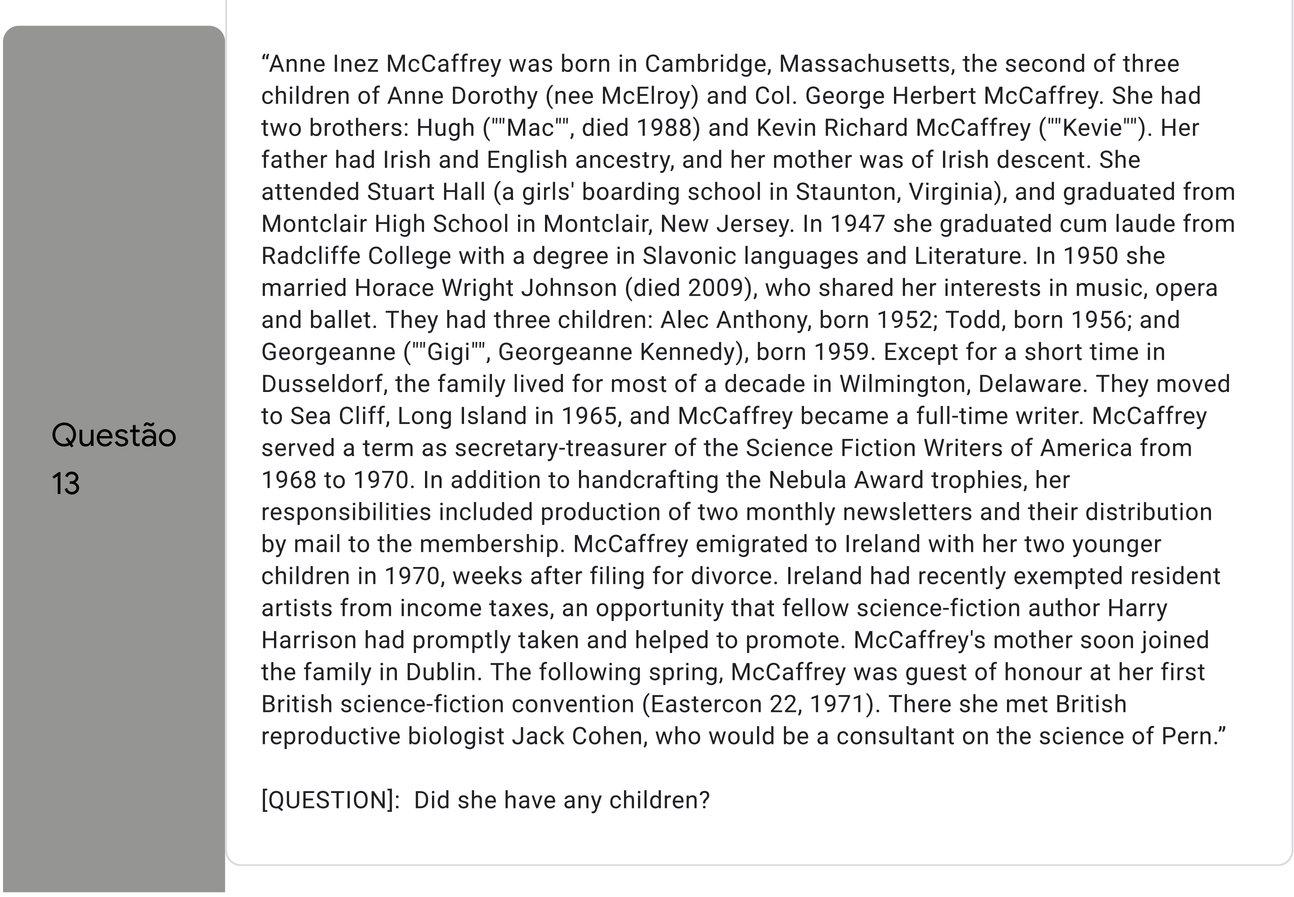
	1	2	3	
Não relacionada	○	○	○	Relação evidente

52. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não



Questão 13

“Anne Inez McCaffrey was born in Cambridge, Massachusetts, the second of three children of Anne Dorothy (nee McElroy) and Col. George Herbert McCaffrey. She had two brothers: Hugh (""Mac"", died 1988) and Kevin Richard McCaffrey (""Kevie""). Her father had Irish and English ancestry, and her mother was of Irish descent. She attended Stuart Hall (a girls' boarding school in Staunton, Virginia), and graduated from Montclair High School in Montclair, New Jersey. In 1947 she graduated cum laude from Radcliffe College with a degree in Slavonic languages and Literature. In 1950 she married Horace Wright Johnson (died 2009), who shared her interests in music, opera and ballet. They had three children: Alec Anthony, born 1952; Todd, born 1956; and Georgeanne (""Gigi"", Georgeanne Kennedy), born 1959. Except for a short time in Dusseldorf, the family lived for most of a decade in Wilmington, Delaware. They moved to Sea Cliff, Long Island in 1965, and McCaffrey became a full-time writer. McCaffrey served a term as secretary-treasurer of the Science Fiction Writers of America from 1968 to 1970. In addition to handcrafting the Nebula Award trophies, her responsibilities included production of two monthly newsletters and their distribution by mail to the membership. McCaffrey emigrated to Ireland with her two younger children in 1970, weeks after filing for divorce. Ireland had recently exempted resident artists from income taxes, an opportunity that fellow science-fiction author Harry Harrison had promptly taken and helped to promote. McCaffrey's mother soon joined the family in Dublin. The following spring, McCaffrey was guest of honour at her first British science-fiction convention (Eastercon 22, 1971). There she met British reproductive biologist Jack Cohen, who would be a consultant on the science of Pern.”

[QUESTION]: Did she have any children?

53. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

54. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

55. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

56. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não

Questão 14

“In 1973, while Zahir Shah was in Italy, undergoing eye surgery and therapy for lumbago, his cousin and former Prime Minister Mohammed Daoud Khan staged a coup d'etat and established a republican government. As a former prime minister, Daoud Khan had been forced to resign by Zahir Shah a decade earlier. During August 1974, Zahir Shah abdicated rather than risk a civil war, ending over 200 years of royal rule in Afghanistan. Zahir Shah lived in exile in Italy for twenty-nine years in a villa in the affluent community of Olgiata on Via Cassia, north of Rome where he spent his time playing golf and chess, as well as tending to his garden. He was prohibited from returning to Afghanistan during the late 1970s by the Soviet-assisted Communist government. In 1983 during the Soviet-Afghan War, Zahir Shah was cautiously involved with plans to develop a government in exile. Ultimately these plans failed because he could not reach a consensus with the powerful Islamist factions. It has also been reported that Afghanistan, the Soviet Union and India had all tried to persuade Zahir Shah to return as chief of a neutral, possibly interim, administration in Kabul. In 1991, Zahir Shah survived an attempt on his life by a knife-wielding assassin masquerading as a Portuguese journalist. After the fall of the pro-Soviet government, Zahir Shah was favored by many to return and restore the monarchy to unify the country and as he was acceptable to most factions. However these efforts were blocked mostly by Pakistan's ISI, who feared his stance on the Durand Line issue. In June 1995, Zahir Shah's former envoy Sardar Wali announced at talks in Islamabad, Pakistan that Zahir Shah was willing to participate in peace talks to end the Afghan Civil War, but no consensus”

[QUESTION]: Did he have any accomplishments during this time ?

57. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

58. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

59. Relevância *

Marcar apenas uma oval.

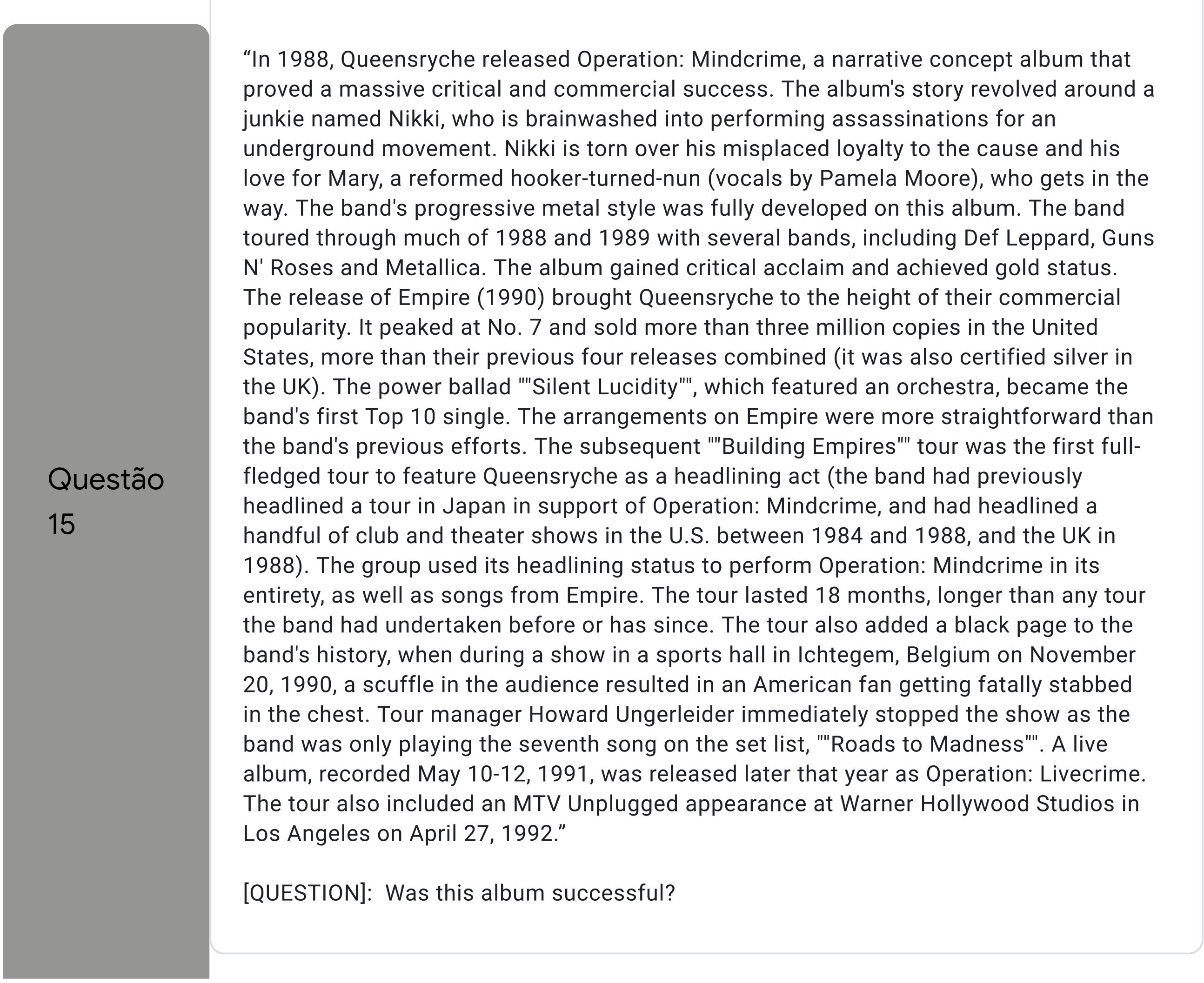
	1	2	3	
Não relacionada	○	○	○	Relação evidente

60. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não



Questão 15

“In 1988, Queensryche released Operation: Mindcrime, a narrative concept album that proved a massive critical and commercial success. The album's story revolved around a junkie named Nikki, who is brainwashed into performing assassinations for an underground movement. Nikki is torn over his misplaced loyalty to the cause and his love for Mary, a reformed hooker-turned-nun (vocals by Pamela Moore), who gets in the way. The band's progressive metal style was fully developed on this album. The band toured through much of 1988 and 1989 with several bands, including Def Leppard, Guns N' Roses and Metallica. The album gained critical acclaim and achieved gold status. The release of Empire (1990) brought Queensryche to the height of their commercial popularity. It peaked at No. 7 and sold more than three million copies in the United States, more than their previous four releases combined (it was also certified silver in the UK). The power ballad ""Silent Lucidity"", which featured an orchestra, became the band's first Top 10 single. The arrangements on Empire were more straightforward than the band's previous efforts. The subsequent ""Building Empires"" tour was the first full-fledged tour to feature Queensryche as a headlining act (the band had previously headlined a tour in Japan in support of Operation: Mindcrime, and had headlined a handful of club and theater shows in the U.S. between 1984 and 1988, and the UK in 1988). The group used its headlining status to perform Operation: Mindcrime in its entirety, as well as songs from Empire. The tour lasted 18 months, longer than any tour the band had undertaken before or has since. The tour also added a black page to the band's history, when during a show in a sports hall in Ichtegem, Belgium on November 20, 1990, a scuffle in the audience resulted in an American fan getting fatally stabbed in the chest. Tour manager Howard Ungerleider immediately stopped the show as the band was only playing the seventh song on the set list, ""Roads to Madness"". A live album, recorded May 10-12, 1991, was released later that year as Operation: Livecrime. The tour also included an MTV Unplugged appearance at Warner Hollywood Studios in Los Angeles on April 27, 1992.”

[QUESTION]: Was this album successful?

61. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

62. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

63. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

64. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não

Questão 16

“The adjective indigenous was historically used to describe animals and plant origins. During the late twentieth century, the term Indigenous people began to be used to describe a legal category in indigenous law created in international and national legislations; it refers to culturally distinct groups affected by colonization. It is derived from the Latin word indigena, which is based on the root gen- 'to be born' with an archaic form of the prefix in 'in'. Any given people, ethnic group or community may be described as indigenous in reference to some particular region or location that they see as their traditional tribal land claim. Other terms used to refer to indigenous populations are aboriginal, native, original, or first (as in Canada's First Nations). The use of the term peoples in association with the indigenous is derived from the 19th century anthropological and ethnographic disciplines that Merriam-Webster Dictionary defines as ""a body of persons that are united by a common culture, tradition, or sense of kinship, which typically have common language, institutions, and beliefs, and often constitute a politically organized group"". James Anaya, former Special Rapporteur on the Rights of Indigenous Peoples, has defined indigenous peoples as ""living descendants of pre-invasion inhabitants of lands now dominated by others. They are culturally distinct groups that find themselves engulfed by other settler societies born of forces of empire and conquest"". They form at present non-dominant sectors of society and are determined to preserve, develop and transmit to future generations their ancestral territories, and their ethnic identity, as the basis of their continued existence as peoples, in accordance with their own cultural patterns, social institutions and legal system. The International Day of the World's Indigenous People falls on 9 August as this was the date of the first meeting in 1982 of the United Nations Working Group of Indigenous Populations of the Subcommission on Prevention of Discrimination and Protection of Minorities of the Commission on Human Rights”

[QUESTION]: What is the word for the word “race”?

65. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

66. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

67. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

68. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não

Questão 17

“The rocks on the plains of Gusev are a type of basalt. They contain the minerals olivine, pyroxene, plagioclase, and magnetite, and they look like volcanic basalt as they are fine-grained with irregular holes (geologists would say they have vesicles and vugs). Much of the soil on the plains came from the breakdown of the local rocks. Fairly high levels of nickel were found in some soils; probably from meteorites. Analysis shows that the rocks have been slightly altered by tiny amounts of water. Outside coatings and cracks inside the rocks suggest water deposited minerals, maybe bromine compounds. All the rocks contain a fine coating of dust and one or more harder rinds of material. One type can be brushed off, while another needed to be ground off by the Rock Abrasion Tool (RAT). There are a variety of rocks in the Columbia Hills, some of which have been altered by water, but not by very much water. The dust in Gusev Crater is the same as dust all around the planet. All the dust was found to be magnetic. Moreover, Spirit found the magnetism was caused by the mineral magnetite, especially magnetite that contained the element titanium. One magnet was able to completely divert all dust hence all Martian dust is thought to be magnetic. The spectra of the dust was similar to spectra of bright, low thermal inertia regions like Tharsis and Arabia that have been detected by orbiting satellites. A thin layer of dust, maybe less than one millimeter thick covers all surfaces. Something in it contains a small amount of chemically bound water.”

[QUESTION]: What kind of radiation is made of?

69. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

70. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

71. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

72. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não

Questão 18

“Albert Stevens Crockett, the hotel's veteran publicist and historian, wrote his first cocktail book ""Old Waldorf Bar Days"" in 1931 during Prohibition and the construction of the current hotel on Park Avenue. It was an homage to the original hotel and its famous bar and clientele. The book contains Crockett's takes on the original hand-written leather-bound book of recipes that was given to him at the time of the closure by bartender Joseph Taylor. This edition was never reprinted. In 1934, Crockett wrote a second book, ""The Old Waldorf Astoria Bar Book"", in response to the repeal of the Volstead Act and the end of the Prohibition era. He edited out most of the text from the first book. Drawing from his experiences as a travel writer, Crockett added nearly 150 more recipes, the bulk of which can be found in the ""Cuban Concoctions"" and ""Jamaican Jollifers"" chapters. These books became reference books on the subject of pre-Prohibition cocktails and its culture. In 2016, the long-time hotel bar manager of Peacock Alley and La Chine, Frank Caiafa, added a completely new edition to the canon. Caiafa's ""The Waldorf Astoria Bar Book"" includes all of the recipes in Crockett's books; many of the hotel's most important recipes created since 1935; and his own creations. In 2017, it was nominated for a James Beard Foundation Award for Best Beverage Book. Other notable books with connections to the hotel include ""Drinks"" (1914) by Jacques Straub, a wine steward and friend of Oscar Tschirky who had written about the first hotel's notable recipes. Tschirky himself compiled a list of 100 recipes for his own book ""100 Famous Cocktails"" (1934), a selection of favorites from Crockett's books. Finally, hotel publicist Ted Saucier wrote ""Bottoms Up"" in 1951, consisting of a compendium of popular, national recipes of the day.”

[QUESTION]: Any other interesting facts about his early career ?

73. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

74. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

75. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

76. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não

Questão 19

“Prior to the 2008 Republican National Convention, a Gallup poll found that most voters were unfamiliar with Sarah Palin. During her campaign to become vice president, 39% said Palin was ready to serve as president if needed, 33% said Palin was not, and 29% had no opinion. This was ""the lowest vote of confidence in a running mate since the elder George Bush chose then-Indiana senator Dan Quayle to join his ticket in 1988."" Following the convention, her image came under close media scrutiny, particularly with regard to her religious perspective on public life, her socially conservative views, and her perceived lack of experience. Palin's experience in foreign and domestic politics came under criticism among conservatives as well as liberals following her nomination. At the same time, Palin became more popular than John McCain among Republicans. One month after McCain announced Palin as his running mate, she was viewed both more favorably and unfavorably among voters than her opponent, Delaware Senator Joe Biden. A plurality of the television audience rated Biden's performance higher at the 2008 vice-presidential debate. Media outlets repeated Palin's statement that she ""stood up to Big Oil"" when she resigned after 11 months as the head of the Alaska Oil and Gas Conservation Commission, due to abuses she witnessed involving other Republican commissioners and their ties to energy companies and energy lobbyists, and again when she raised taxes on oil companies as governor. In turn, others have said that Palin is a ""friend of Big Oil"" due to her advocacy of oil exploration and development including drilling in the Arctic National Wildlife Refuge and the de-listing of the polar bear as an endangered species. Palin was named one of America's ""10 Most Fascinating People of 2008"" by Barbara Walters for an ABC special on December 4, 2008. In April 2010, she was selected as one of the world's 100 most influential people by TIME Magazine.”

[QUESTION]: Are there any other interesting aspects about this article?

77. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

78. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

79. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

80. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não

Questão 20

Barry Humphries was invited to join the fledgling Union Theatre Repertory Company early in 1955 and toured Victorian country towns performing Twelfth Night, directed by Ray Lawler. On tour, Humphries invented Edna gradually as part of the entertainment for the actors during commutes between country towns. Humphries gradually developed a falsetto impersonation of a Melbourne housewife, imitating the Country Women's Association representatives who welcomed the troupe in each town. At Lawler's suggestion, Mrs Everage (later named Edna after Humphries' nanny) made her first appearance in a Melbourne University's UTRC revue at the end of 1955, as the city prepared for the 1956 Olympic Games. The sketch involved a houseproud ""average housewife"" offering her Moonee Ponds home as an Olympic billet, spruiking her home as possessing ""burgundy wall-to-wall carpets, lamington cakes and reindeers frosted on glass dining-room doors"". At this time the character was billed as ""Mrs Norm Everage"" (Humphries describing this name as ""Everage as in 'average', husband Norm as in 'normal'"") and had none of the characteristic flamboyant wardrobe of later years. His mother (whom the interviewer William Cook said ""sounds like a frightful snob"") was a major inspiration for Edna, although he denied it when she was alive to protect her feelings. Her first monologue in 1955 was about her ""lovely home"", reflecting young Barry's own site visits accompanying his builder father. Originally she was a ""mousy"" character and too quiet to please the raucous crowd at The Establishment club in London. According to one author, Edna came into her own during the 1980s when the policies of Thatcherism--and what he described as the ""vindictive style of the times""--allowed Dame Edna to sharpen her observations accordingly. Lahr wrote that Edna took Prime Minister Margaret Thatcher's ""seemingly hypocritical motto"" of ""caring and compassion"" for others and turned it on its head, Edna became the voice of Humphries' outrage.

[QUESTION]: Did he win any awards ?

81. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

82. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

83. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

84. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim
○ Não

Questão 21

“Another important library the University Library, founded in 1816, is home to over two million items. The building was designed by architects Marek Budzyaski and Zbigniew Badowski and opened on 15 December 1999. It is surrounded by green. The University Library garden, designed by Irena Bajerska, was opened on 12 June 2002. It is one of the largest and most beautiful roof gardens in Europe with an area of more than 10,000 m2 (107,639.10 sq ft), and plants covering 5,111 m2 (55,014.35 sq ft). As the university garden it is open to the public every day."

[QUESTION]: What did they do at the royal museum?

85. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

86. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

87. Relevância *

Marcar apenas uma oval.

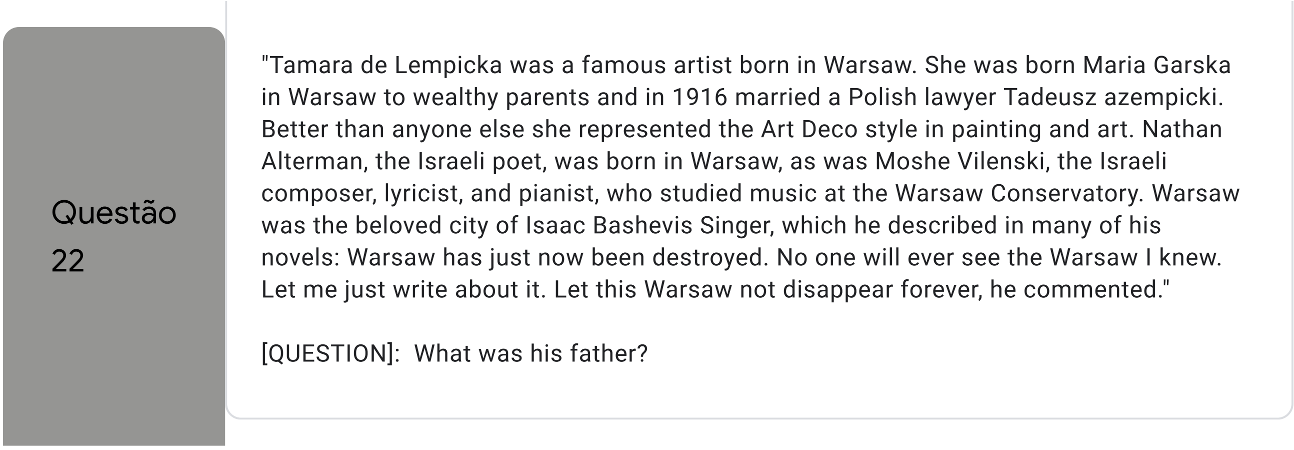
	1	2	3	
Não relacionada	○	○	○	Relação evidente

88. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não



Questão 22

"Tamara de Lempicka was a famous artist born in Warsaw. She was born Maria Garska in Warsaw to wealthy parents and in 1916 married a Polish lawyer Tadeusz azempicki. Better than anyone else she represented the Art Deco style in painting and art. Nathan Alterman, the Israeli poet, was born in Warsaw, as was Moshe Vilenski, the Israeli composer, lyricist, and pianist, who studied music at the Warsaw Conservatory. Warsaw was the beloved city of Isaac Bashevis Singer, which he described in many of his novels: Warsaw has just now been destroyed. No one will ever see the Warsaw I knew. Let me just write about it. Let this Warsaw not disappear forever, he commented."

[QUESTION]: What was his father?

89. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

90. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

91. Relevância *

Marcar apenas uma oval.

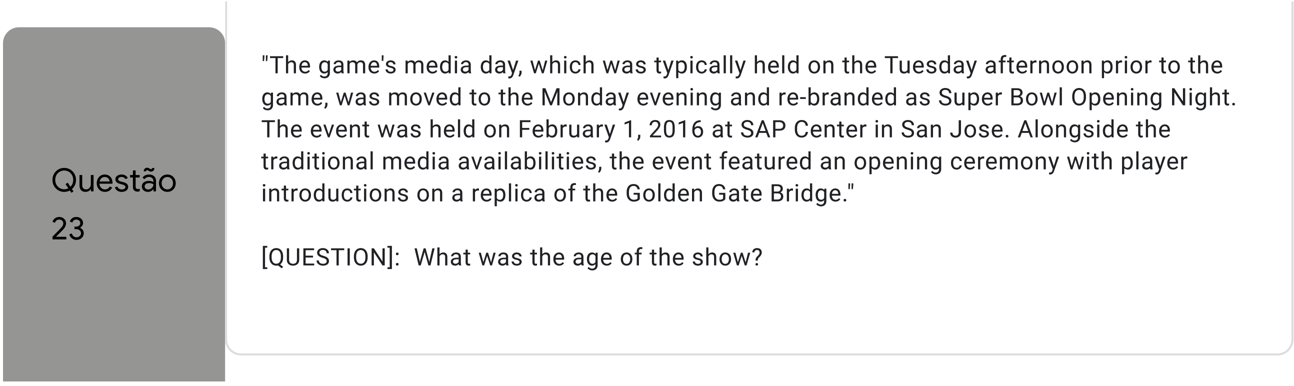
	1	2	3	
Não relacionada	○	○	○	Relação evidente

92. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não



Questão 23

"The game's media day, which was typically held on the Tuesday afternoon prior to the game, was moved to the Monday evening and re-branded as Super Bowl Opening Night. The event was held on February 1, 2016 at SAP Center in San Jose. Alongside the traditional media availabilities, the event featured an opening ceremony with player introductions on a replica of the Golden Gate Bridge."

[QUESTION]: What was the age of the show?

93. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

94. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

95. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

96. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não

Questão 24

"Carolina suffered a major setback when Thomas Davis, an 11-year veteran who had already overcome three ACL tears in his career, went down with a broken arm in the NFC Championship Game. Despite this, he insisted he would still find a way to play in the Super Bowl. His prediction turned out to be accurate."

[QUESTION]: Why did they do in the early years?

97. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

98. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

99. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

100. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não

Questão 25

"As the designated home team in the annual rotation between AFC and NFC teams, the Broncos elected to wear their road white jerseys with matching white pants. Elway stated, ""We've had Super Bowl success in our white uniforms."" The Broncos last wore matching white jerseys and pants in the Super Bowl in Super Bowl XXXIII, Elway's last game as Denver QB, when they defeated the Atlanta Falcons 34â€“19. In their only other Super Bowl win in Super Bowl XXXII, Denver wore blue jerseys, which was their primary color at the time. They also lost Super Bowl XXI when they wore white jerseys, but they are 0-4 in Super Bowls when wearing orange jerseys, losing in Super Bowl XII, XXII, XXIV, and XLVIII. The only other AFC champion team to have worn white as the designated home team in the Super Bowl was the Pittsburgh Steelers; they defeated the Seattle Seahawks 21-10 in Super Bowl XL 10 seasons prior. The Broncos' decision to wear white meant the Panthers would wear their standard home uniform: black jerseys with silver pants."

[QUESTION]: How did they do in the semi-finals?

101. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

102. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

103. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

104. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não

Questão 26

"In 1874, Tesla evaded being drafted into the Austro-Hungarian Army in Smiljan by running away to Tomingaj, near Grazac. There, he explored the mountains in hunter's garb. Tesla said that this contact with nature made him stronger, both physically and mentally. He read many books while in Tomingaj, and later said that Mark Twain's works had helped him to miraculously recover from his earlier illness."

[QUESTION]: Why did he approach rogers?

105. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

106. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

107. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

108. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não

Questão 27

“The origin of the legendary figure is not fully known. The best-known legend, by Artur Oppman, is that long ago two of Triton's daughters set out on a journey through the depths of the oceans and seas. One of them decided to stay on the coast of Denmark and can be seen sitting at the entrance to the port of Copenhagen. The second mermaid reached the mouth of the Vistula River and plunged into its waters. She stopped to rest on a sandy beach by the village of Warszowa, where fishermen came to admire her beauty and listen to her beautiful voice. A greedy merchant also heard her songs; he followed the fishermen and captured the mermaid.”

[QUESTION]: Who is the only of the apartment?

109. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

110. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

111. Relevância *

Marcar apenas uma oval.

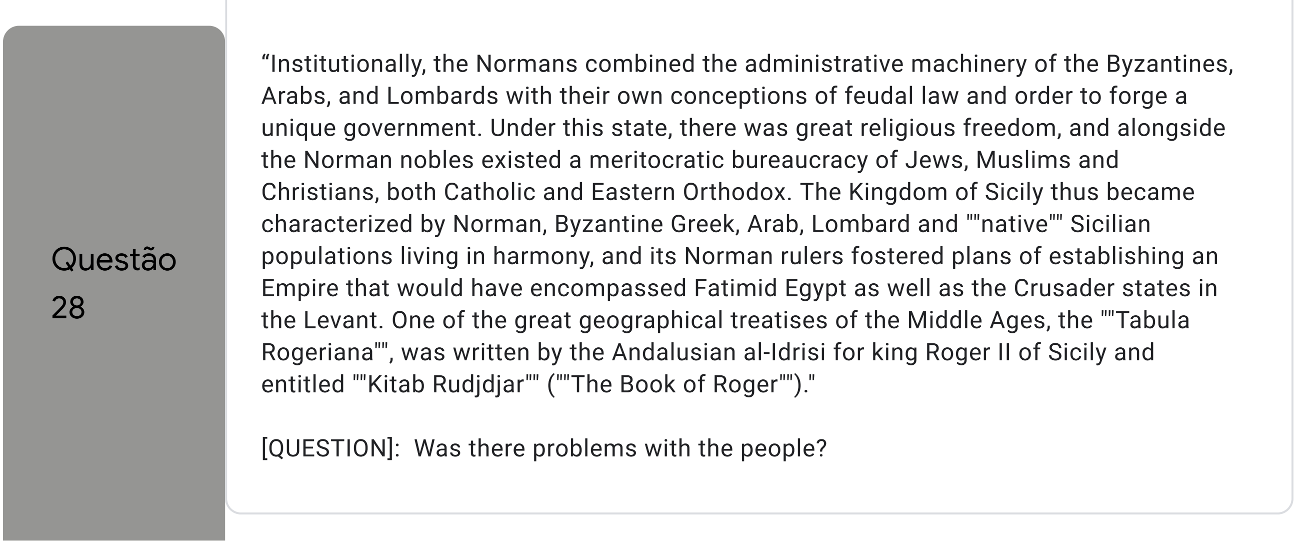
	1	2	3	
Não relacionada	○	○	○	Relação evidente

112. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não



Questão 28

“Institutionally, the Normans combined the administrative machinery of the Byzantines, Arabs, and Lombards with their own conceptions of feudal law and order to forge a unique government. Under this state, there was great religious freedom, and alongside the Norman nobles existed a meritocratic bureaucracy of Jews, Muslims and Christians, both Catholic and Eastern Orthodox. The Kingdom of Sicily thus became characterized by Norman, Byzantine Greek, Arab, Lombard and ""native"" Sicilian populations living in harmony, and its Norman rulers fostered plans of establishing an Empire that would have encompassed Fatimid Egypt as well as the Crusader states in the Levant. One of the great geographical treatises of the Middle Ages, the ""Tabula Rogeriana"", was written by the Andalusian al-Idrisi for king Roger II of Sicily and entitled ""Kitab Rudjdjar"" (""The Book of Roger"")."

[QUESTION]: Was there problems with the people?

113. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

114. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

115. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

116. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não

Questão 29

“In April 1191 Richard the Lion-hearted left Messina with a large fleet in order to reach Acre. But a storm dispersed the fleet. After some searching, it was discovered that the boat carrying his sister and his fiancé Berengaria was anchored on the south coast of Cyprus, together with the wrecks of several other ships, including the treasure ship. Survivors of the wrecks had been taken prisoner by the island's despot Isaac Komnenos. On 1 May 1191, Richard's fleet arrived in the port of Limassol on Cyprus. He ordered Isaac to release the prisoners and the treasure. Isaac refused, so Richard landed his troops and took Limassol."

[QUESTION]: How many ships did the fleet fleet ?

117. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

118. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

119. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

120. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não

Questão 30

“In 1873, Tesla returned to his birthtown, Smiljan. Shortly after he arrived, Tesla contracted cholera; he was bedridden for nine months and was near death multiple times. Tesla's father, in a moment of despair, promised to send him to the best engineering school if he recovered from the illness (his father had originally wanted him to enter the priesthood)."

[QUESTION]: How long did he stay there ?

121. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

122. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

123. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

124. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não

Questão 31

Between Bingen and Bonn, the Middle Rhine flows through the Rhine Gorge, a formation which was created by erosion. The rate of erosion equaled the uplift in the region, such that the river was left at about its original level while the surrounding lands raised. The gorge is quite deep and is the stretch of the river which is known for its many castles and vineyards. It is a UNESCO World Heritage Site (2002) and known as ""the Romantic Rhine"", with more than 40 castles and fortresses from the Middle Ages and many quaint and lovely country villages."

[QUESTION]: What be he for ?

125. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

126. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

127. Relevância *

Marcar apenas uma oval.

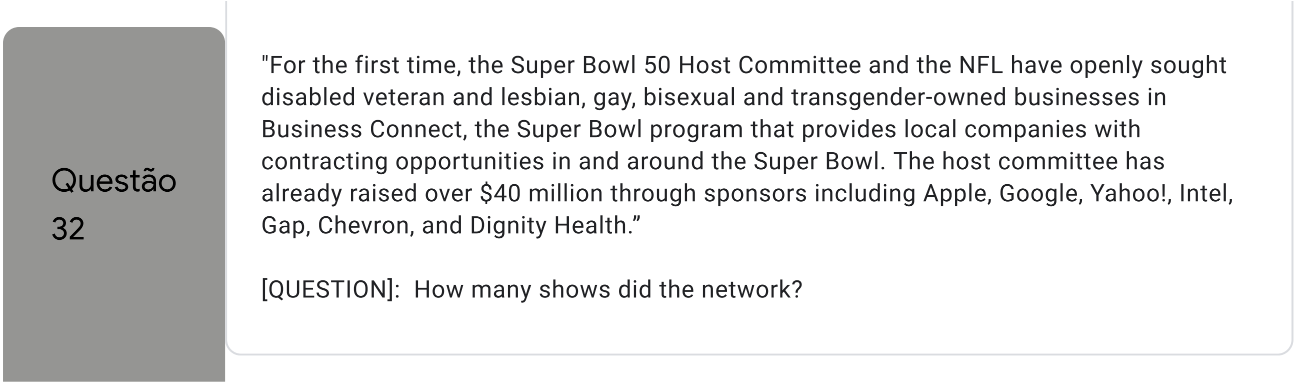
	1	2	3	
Não relacionada	○	○	○	Relação evidente

128. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não



Questão 32

"For the first time, the Super Bowl 50 Host Committee and the NFL have openly sought disabled veteran and lesbian, gay, bisexual and transgender-owned businesses in Business Connect, the Super Bowl program that provides local companies with contracting opportunities in and around the Super Bowl. The host committee has already raised over $40 million through sponsors including Apple, Google, Yahoo!, Intel, Gap, Chevron, and Dignity Health."

[QUESTION]: How many shows did the network?

129. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

130. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

131. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

132. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não

Questão 33

"QuickBooks sponsored a "Small Business Big Game" contest, in which Death Wish Coffee had a 30-second commercial aired free of charge courtesy of QuickBooks. Death Wish Coffee beat out nine other contenders from across the United States for the free advertisement."

[QUESTION]: How many years did the contract last?

133. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

134. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

135. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

136. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não

Questão 34

“In 1875, Tesla enrolled at Austrian Polytechnic in Graz, Austria, on a Military Frontier scholarship. During his first year, Tesla never missed a lecture, earned the highest grades possible, passed nine exams (nearly twice as many required), started a Serbian culture club, and even received a letter of commendation from the dean of the technical faculty to his father, which stated, ""Your son is a star of first rank."" Tesla claimed that he worked from 3 a.m. to 11 p.m., no Sundays or holidays excepted. He was ""mortified when [his] father made light of [those] hard won honors."" After his father's death in 1879, Tesla found a package of letters from his professors to his father, warning that unless he were removed from the school, Tesla would be killed through overwork. During his second year, Tesla came into conflict with Professor Poeschl over the Gramme dynamo, when Tesla suggested that commutators weren't necessary. At the end of his second year, Tesla lost his scholarship and became addicted to gambling. During his third year, Tesla gambled away his allowance and his tuition money, later gambling back his initial losses and returning the balance to his family. Tesla said that he ""conquered [his] passion then and there,"" but later he was known to play billiards in the US. When exam time came, Tesla was unprepared and asked for an extension to study but was denied. He never graduated from the university and did not receive grades for the last semester."

[QUESTION]: How long did he stay a student?

137. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

138. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

139. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

140. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não

Questão 35

“One of the most famous people born in Warsaw was Maria Curie, who achieved international recognition for her research on radioactivity and was the first female recipient of the Nobel Prize. Famous musicians include Szpilman and Frederic Chopin. Though Chopin was born in the village of Åelazowa Wola, about 60 km (37 mi) from Warsaw, he moved to the city with his family when he was seven months old. Casimir Pulaski, a Polish general and hero of the American Revolutionary War, was born here in 1745."

[QUESTION]: Did he have any family?

141. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

142. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

143. Relevância *

Marcar apenas uma oval.

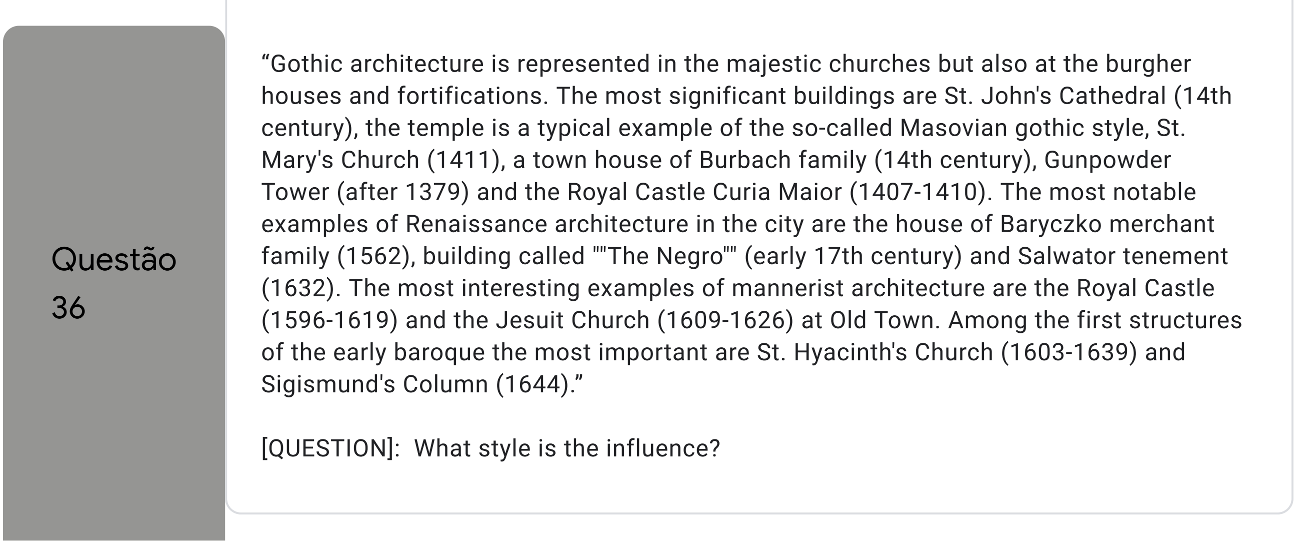
	1	2	3	
Não relacionada	○	○	○	Relação evidente

144. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não



Questão 36

"Gothic architecture is represented in the majestic churches but also at the burgher houses and fortifications. The most significant buildings are St. John's Cathedral (14th century), the temple is a typical example of the so-called Masovian gothic style, St. Mary's Church (1411), a town house of Burbach family (14th century), Gunpowder Tower (after 1379) and the Royal Castle Curia Maior (1407-1410). The most notable examples of Renaissance architecture in the city are the house of Baryczko merchant family (1562), building called ""The Negro"" (early 17th century) and Salwator tenement (1632). The most interesting examples of mannerist architecture are the Royal Castle (1596-1619) and the Jesuit Church (1609-1626) at Old Town. Among the first structures of the early baroque the most important are St. Hyacinth's Church (1603-1639) and Sigismund's Column (1644)."

[QUESTION]: What style is the influence?

145. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

146. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

147. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

148. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim
○ Não

Questão 37

"CBS provided digital streams of the game via CBSSports.com, and the CBS Sports apps on tablets, Windows 10, Xbox One and other digital media players (such as Chromecast and Roku). Due to Verizon Communications exclusivity, streaming on smartphones was only provided to Verizon Wireless customers via the NFL Mobile service. The ESPN Deportes Spanish broadcast was made available through WatchESPN."

[QUESTION]: Who are these major networks?

149. Bem escrita ? *

Marcar apenas uma oval.

	1	2	3	
Erros	○	○	○	Sem erros

150. Compreensibilidade *

Marcar apenas uma oval.

	1	2	3	
Incompreensível	○	○	○	Compreensível

151. Relevância *

Marcar apenas uma oval.

	1	2	3	
Não relacionada	○	○	○	Relação evidente

152. Tem resposta no texto ? *

Marcar apenas uma oval.

○ Sim

○ Não

Este conteúdo não foi criado nem aprovado pelo Google.

Google Formulários