# KNOWING THE RESEARCH ON THE USE OF APPROACHES RELATED TO PROBABILISTIC LANGUAGE IN THE EARLY YEARS OF ELEMENTARY SCHOOL IN BRAZIL

# CONHECENDO A PESQUISA SOBRE A UTILIZAÇÃO DE ABORDAGENS REFERENTES À LINGUAGEM PROBABILÍSTICA NOS ANOS INICIAIS DO ENSINO FUNDAMENTAL NO BRASIL

# CONOCIENDO LA INVESTIGACIÓN SOBRE EL USO DE ENFOQUES RELACIONADOS CON EL LENGUAJE PROBABILÍSTICO EN LOS PRIMEROS AÑOS DE LA ENSEÑANZA BÁSICA EN BRASIL

OLIVEIRA JÚNIOR, Ailton Paulo de
ailton.junior@ufabc.edu.br
UFABC – Universidade Federal do ABC
https://orcid.org/0000-0002-2721-7192

KIAN, Fátima Aparecida
fatima.kian@ufabc.edu.br
UFABC – Universidade Federal do ABC
https://orcid.org/0000-0003-0105-7335

SILVA SANTOS, Luzia Roseli da
luzia.santos@ufabc.edu.br
UFABC – Universidade Federal do ABC
https://orcid.org/0000-0001-6930-9215

**ABSTRACT** The objective of this study was to systematically analyze Brazilian literature through scientific articles, dissertations and theses that used approaches for the assessment of probabilistic language in the early years of elementary school. Seven studies were selected, covering the period from 2007 to 2021, and the IRaMuTeQ software was used for multivariate textual analysis, through organization into texts. The results indicate that while most research focuses on students in the early years of elementary school and in initial teacher training (Pedagogy), there is rather a lack of studies that focus on teachers in effective exercise on their profession. We emphasize that this research field is still incipient, evidencing concerns regarding students' profile and their relationships with experiential knowledge and everyday life.
**Keywords**: Elementary school. Multivariate textual analysis. Probabilistic language. Systematic literature review.

**RESUMO** O objetivo deste trabalho foi analisar sistematicamente a literatura brasileira por meio de artigos científicos, dissertações e teses que utilizaram abordagens para a avaliação da linguagem probabilística nos anos iniciais do ensino fundamental. Sete estudos foram selecionados, no período de 2007 até 2021, e o *software* IRaMuTeQ foi utilizado para a análise textual multivariada, por meio da organização em textos. Os resultados indicam que enquanto a maioria das pesquisas é voltada aos alunos dos anos iniciais do ensino fundamental e em formação inicial de professores (Pedagogia), há uma lacuna de trabalhos mais voltados aos professores em efetivo exercício de sua profissão. Reforçamos que as pesquisas ainda são incipientes, evidenciando a preocupação dessas com o perfil do aluno e suas relações com o conhecimento experiencial e com o cotidiano.
**Palavras-chave**: Análise textual multivariada. Ensino fundamental. Linguagem probabilística. Revisão sistemática de literatura.


**RESUMEN** El objetivo de este trabajo fue analizar sistemáticamente la literatura brasileña a través de artículos científicos, disertaciones y tesis que utilizaron enfoques para la evaluación del lenguaje probabilístico en los primeros años de la escuela primaria. Se seleccionaron siete estudios, de 2007 a 2021, y se utilizó el software IRaMuTeQ para el análisis textual multivariado, a través de la organización en textos. Los resultados indican que mientras la mayor parte de la investigación está dirigida a estudiantes de los primeros años de la enseñanza primaria y en formación inicial del profesorado (Pedagogía), hay una lacuna de trabajos más dirigidos a docentes en ejercicio efectivo de su profesión. Reforzamos que las investigaciones aún son incipientes, evidenciando su preocupación con el perfil del estudiante y sus relaciones con el conocimiento experiencial y la vida cotidiana.
**Palabras Clave**: Análisis textual multivariado. Enseñanza primaria. Lenguaje probabilístico. Revisión sistemática de la literatura.

## 1 INTRODUCTION

The study of knowledge of probabilistic concepts in the early years of Brazilian elementary school (6 to 10 years old) is essential to create strategies for children's learning. This is evident since, in today's world, we receive a large amount of information daily, requiring an understanding of random or non-random phenomena.

In addition, according to the National Common Curricular Base - BNCC (Brasil, 2018), the development of concepts of a probabilistic nature should be encouraged from the early years of Elementary School.

We believe that it is necessary to research what the ideas expressed by children and adults in their daily lives are through their own words. It is essential to promote a

conceptual understanding of probabilistic terms, since, according to Van Dooren et al. (2003), students have misconceptions towards these concepts.

Based on Findley, Whitacre and Atabas (2020), we reinforce the importance of understanding knowledge and use of everyday terms to later subsidize probabilistic concepts and terms. Furthermore, by understanding individual experiences and their informal naming practices, this knowledge can be established to be brought into the curricular context and into broader instructional practice.

In addition, for Vygotsky (2001), the meaning of a word is a phenomenon of thought to the extent that it is incorporated into speech. Speech is meaningful only to the extent that it is connected with thought and illuminated by it, as thought is a dynamic process that is incorporated into speech.

We agree with Góes and Cruz (2006) that, although the meaning of the word is an act of generalization, it changes as it encounters new situations of its use and as its intellectual processes of abstraction and generalization are developed.

Thus, until the shared meaning among members of a given social environment is understood, the meaning of a word changes for them. Therefore, the objective of this study was to systematically analyze academic works aimed at describing and identifying approaches for the assessment of probabilistic language in the early years of elementary education in Brazil.

## 2 THEORETICAL FRAMEWORK

Regarding probabilistic language, Vásquez (2018) argues that it has a remarkably close connection with everyday language, claiming that the first linguistic elements that are part of the language end up being a fundamental element for research related to teaching probability, mainly in the early ages of basic education.

Furthermore, Gal (2005) highlights that probabilistic literacy encompasses cognitive elements, such as probabilistic calculations, language, contexts and critical questions, as well as dispositional elements which include beliefs, attitudes and habits. Consequently, there is a need for people to go through probabilistic literacy to deal with the wide range of real-world situations involving the generation or interpretation of probabilistic messages, as well as decision-making about phenomena that appear.

Vásquez (2018) also emphasizes that, in order to provide students with the tools they need to develop probabilistic literacy more effectively, they must access, use, interpret, and communicate information and ideas related to probability. Additionally, Vásquez maintains that the teaching of probability must be conducted continuously, beginning in early education and continuing through high school. Thus, students will be able to develop knowledge that allows a critical way of thinking and a better understanding of situations present in daily life, expressed as random phenomena, chance and uncertainty.

Regarding probabilistic language, specifically in elementary school textbooks, Torres et al. (2013) highlight the richness and diversity of verbal expressions and colloquial language that prevails over formal language. They also indicate that language is linked to various meanings of probability, whether intuitive, classical, frequencial and subjective. They also observe the numerical language that develops according to the introduction of different numerical systems in teaching.

They also show that current curriculum documents in Spain suggest that children should acquire elementary probabilistic language through games, experiments and observation of natural phenomena. So that they can then identify random situations until they understand the determination of simple probability. Thus, a fundamental element for the construction of mathematical knowledge is the language used through child's everyday language, which progresses until it becomes a language with a higher level of abstraction (Torres et al., 2013).

Vásquez (2014) explains that, due to the connection between everyday expressions and probabilistic language, the first linguistic elements that support the natural language spoken in everyday life end up being an essential element for probabilistic studies. It is important for students to receive experiences that help them appreciate the precise power of this language, preventing the student from being prematurely forced into formal mathematical language.

Thus, in the sections that follow, we address the methodology used in this study, present our systematic review on the subject and, finally, the research trend, its advances and weaknesses.

## 3 METHODOLOGY

The study was conducted under the guidelines proposed by Kitchenham et al. (2007) to conduct a Systematic Literature Review - SLR, comprising three phases: planning, process and results reporting. In the planning phase, a protocol for the literature review was elaborated, establishing: 1) the interaction that researchers should have; 2) defining the procedure for conducting the review; 3) the formulation of the research question; 4) indication of search strategies, inclusion and exclusion criteria, data collection and analysis.

The studies were categorized in order to answer the research question: How has research employing didactic approaches to probabilistic language in the early years of elementary school contributed to probability teaching in Brazil?

The query was conducted online, using the main national databases, specifically in this order: (1) Public Domain Portal; (2) SciELO - Scientific Electronic Library Online; (3) Brazilian Digital Library of Theses and Dissertations (BDTD); (4) Catalog of Theses and Dissertations of the Coordination for the Improvement of Higher Education Personnel - CAPES; (5) Google Scholar.
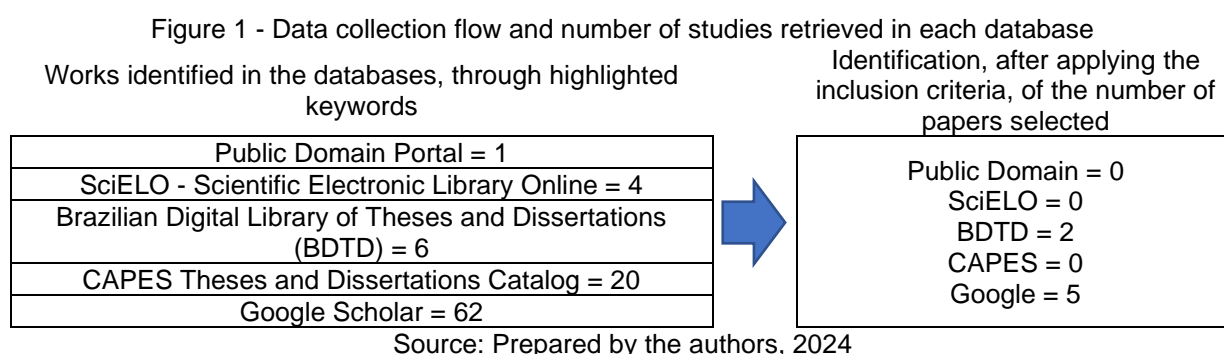
The following keywords were used: Language as the first option (E) Probability (E) Teaching (E) Education (E) Elementary School (E) Final years. The criteria adopted to compose the corpus were: (a) theses or dissertations published in Brazil; (b) articles published in Brazilian scientific journals; (c) books published in Brazil; (d) national and international scientific events by Brazilian researchers.

The inclusion criterion adopted was linguistic, that is, in an overview of academic production in Portuguese (privileging Brazilian publications), which mentioned in the title or abstract the identification of approaches for the assessment of probabilistic language in the early years of elementary school in Brazil. Exclusion criteria included those studies that did not align with the theme, were duplicates, or were published in other languages. Full-text readings were conducted to finalize the selection. Still in Kitchenham et al. (2007), in the second phase, the process focused on the execution of the review protocol. Specifically, we conducted a search in the databases described above using the keyword system, by reading the title and abstract of each work.

Once the search was completed, we evaluated its results and selected eligible studies. To avoid bias in the selection of studies, this was conducted by two

researchers, performing the selection independently and based on the eligibility criteria of the review. Each reviewer recorded whether or not they agreed with the inclusion of the study, based on the evaluation of titles, abstracts and full texts, in that order. Discordant cases were resolved by consensus or through a third researcher.

Figure 1 shows the flow of data collection, with the first column presenting the number of papers identified in the databases considered for this research, based on the indicated keywords, and the second column those that were selected after applying the inclusion criteria.

Figure 1 - Data collection flow and number of studies retrieved in each database

| Works identified in the databases, through highlighted keywords | Identification, after applying the inclusion criteria, of the number of papers selected |
|---|---|
| Public Domain Portal = 1 <br> SciELO - Scientific Electronic Library Online = 4 <br> Brazilian Digital Library of Theses and Dissertations (BDTD) = 6 <br> CAPES Theses and Dissertations Catalog = 20 <br> Google Scholar = 62 | Public Domain = 0 <br> SciELO = 0 <br> BDTD = 2 <br> CAPES = 0 <br> Google = 5 |

Source: Prepared by the authors, 2024

Therefore, this search returned seven results and all titles, abstracts and the text were read to identify proposals using approaches to assess probabilistic language in the early years of elementary school in Brazil (Table 1).

Table 1 - Identification of texts selected in the database search.

| Text | Author (year of publication) | Publication type | Educational institution | Text title | Work focus | Audience for which the work is intended |
|---|---|---|---|---|---|---|
| 1 | Amaral (2007) | Professional master's in mathematics | Pontifical Catholic University of São Paulo (PUC/SP) | Statistics and initial training with students of a pedagogy course: reflections on a didactic sequence. | Teaching Statistics and Probability | Higher Education Students |
| 2 | Silva (2016) | Academic master's in mathematics education | Federal University of Pernambuco (UFPE/PE) | It's the coin that says, it's not people who don't: probabilistic knowledge of children in game situations. | Teaching Probability | Elementary School Students |
| 3 | Cavalcanti and Guimarães (2018) | Scientific article | Federal University of Pernambuco (UFPE/PE) | Understandings demonstrated by elementary school students when raising hypotheses, analyzing real data and making decisions. | Teaching Probability | Elementary School Students |
| 4 | Assis (2018) | Academic master's in teaching and science and mathematics | State University of Paraíba (UEPB/PB) | Probability teaching: analysis of a proposal for the final years of elementary school. | Teaching Probability | Elementary School Students |
| 5 | Paim (2019) | Academic master's degree in | Federal University of São Carlos | The state of the art of Brazilian research on statistical and probabilistic literacy. | Teaching Statistics | Teachers in initial and |

| | | teaching Exact Sciences | (UFSCar/SP) | | and Probability | continuing training |
|---|---|---|---|---|---|---|
| 6 | Cavalcanti (2019) | Doctorate in Mathematics and Technological Education | Federal University of Pernambuco (UFPE/PE) | Elementary school students learning about hypotheses, data analysis and conclusions from statistical data. | Teaching Probability | Elementary School Students |
| 7 | Kian (2021) | Academic master's degree in teaching and history of science and mathematics | Federal University of ABC (UFABC/SP) | Probabilistic language at the end of the initial years of elementary school: a path for the development of probabilistic literacy. | Teaching Probability | Elementary School Students |

Source: Prepared by the authors, 2024

Following Kitchenham et al. (2007), in the third phase, based on the results, a final report was prepared. The analysis of data obtained through the search for articles (scientific journals and scientific events), books and Brazilian theses (master and doctor), in the previously listed databases was lexical.

To this end, we used the IraMuTeQ (R Interface for Multidimensional Text and Questionnaire Analysis) software seeking to improve the research work through the optimization of the organization process and the more specific delimitation of the selected texts. This procedure allowed the survey of constituent elements of socially shared representations that highlight traces of mental worlds through lexical worlds and, later, inferred through the content analysis technique (Mutombo, 2013).

The IRaMuTeQ program works with initial context units (UCIs) that can be structured in diverse ways depending on the character of the collected data. When working with the selected works, each text must compose a UCI. The set of UCIs is composed of the corpus of analysis that the program divides into text segments, which are the elementary context units (ECUs).

Specific questions (QE) were proposed at the UCIs, which collect, organize and present relevant information about the development of research aimed at the use of historical approaches in teaching statistics in Brazil, namely: QE1: What was emphasized? QE2: What is the methodology or methodological approach used? QE3: What is the context in which it is developed? QE4: What are the types of studies and areas involved? QE5: What are the main results and conclusions?

Next, we performed a Descending Hierarchical Classification (DHC) in order to give rise to lexical classes characterized by vocabulary and text segments that share the same vocabulary (Camargo; Justo, 2013). In this sense, the different classes that

emerge from the text corpus represent the space of meaning of the narrated words and may suggest elements belonging to studies on how historical approaches are conducted in the teaching of Statistics in Brazil.

The set of each selected work was organized into a single text (corpus), each of which was defined by the IRaMuTeQ program as a text segment. The corpus was organized by command lines called asterisk lines, in which the text identification numbers are informed, followed by indispensable variables for conducting the textual analysis. In this research, the variables were coded as follows:

(1) Text: text_01 and so on up to text_7.

(2) Type of publication: pubType_01, theses and dissertations; pubType_02, articles published in scientific journals.

(3) Year of publication of the text: yearPub_01, published in 2007; yearPub_02, published in 2016; yearPub_03, published in 2018; yearPub_04, published in 2019; yearPub_05, published in 2021.

(4) Audience for which the work is intended: publicTeaching_01, student in the early years of elementary school; publicTeaching_01, student in the early years of elementary school and in the Pedagogy course; publicTeaching_03, teacher.

(5) Work focus: workFocus_01, probability teaching; workFocus_02, teaching statistics and probability.

(6) Brazilian region where the study was developed: PubRegion_01, Northeast region; PubRegion_02, Southeast region.

In addition, the material that make up the textual corpus were configured as defined in the IRaMuTeQ tutorial (Camargo; Justo, 2013), regarding accentuation, use of special characters and formatting. The procedure for organizing the command lines, for the insertion of scientific productions, can be seen in the example concerning the fragment of the second text, in the analyzes considered in Portuguese (here translated into English):

**** *text_02 *pubType_01 *yearPub_02 *cycle *publicTeaching_01 *workFocus_01 * PubRegion_01

Based on the analysis of the report Children's understanding of probability, produced by Bryant and Nunes (2012), the initial design of the present study emerged. For these authors, probability is a complex concept that involves the development of four cognitive requirements.

Thus, we use Reinert's method, which proposes a DHC according to the method described by Reinert (1998), which aims to obtain classes of text segments (TS) that, at the same time, have similar vocabulary among themselves and different vocabulary from the TS of other classes.

We emphasize that the choice to use one or another analysis technique depends on the characteristics of the problem and the research objectives (Leblanc, 2015). In this sense, the researcher's theoretical-methodological framework, plus the support of lexicometric analysis software, can provide greater reliability to the inferences made in qualitative research (Justo; Camargo, 2014; Salvador et al., 2018).

Taking this into account, the presented study describes and discusses the characteristics of the use of IRaMuTeQ in the analysis of data from scientific works focused on the use of probabilistic language in the early years of elementary school published between 2007 and 2021 in Brazil.

Therefore, this analysis is based on lexical proximity and the idea that words used in similar context are associated with the same lexical world and are part of specific mental worlds or representational systems. In this analysis, the text segments are classified according to their respective vocabulary and the set of terms is partitioned according to the frequency of the roots of the words. The system looks for classes formed by words that are significantly associated with that class (significance starts with the chi-square test - $\chi^2$).

According to Oliveira (2015), the chi-square test is one of the most important analyzes of IRaMuTeQ, since the software uses correlation logic, starting from segmentations of the textual corpus, together with the list of reduced forms and the dictionary (in this case in Portuguese) available in it to present a hierarchical scheme of classes. In this way, the text is processed so that vocabulary classes can be identified, making it possible to infer which ideas the textual corpus wants to convey - that is, the analysis is made from statistical logic processed by computer and applied lexically.

## 4 RESULTS AND DISCUSSIONS

Starting the study, the first analysis option that IRaMuTeQ makes available is related to the statistical data of the textual corpus (Figure 2).

Figure 2 - Result of Classification by Reinert's Method: textual statistics

```
+-+-+-+-+-+-+-+-+
|i|R|a|M|u|T|e|Q| - Thu Jul 22 17:24:45 2022
+-+-+-+-+-+-+-+-+
Number of texts: 7
Number of text segments: 129
Number of forms: 1270
Number of occurrences: 4544
number of lemmas: 917
Number of active forms: 802
Number of supplementary forms: 105
Number of active forms with frequency >= 3: 250
Average shapes per segment: 35.224806
Number of clusters: 5
92 segments classified on 129 (81.32%)
```

Source: IraMuTeQ software output, 2023

The result of the analysis of textual statistics brings information that summarizes the textual corpus as follows:

a) Number of texts: Number of texts (records) contained in the corpus. In this case, for example, the corpus has seven texts (described in Table 1), corresponding to the paragraphs indicating the specific questions (SQ), which, in the case of this work, collect, organize and present relevant information about the development of research. Focused on the use of probabilistic language for the early years of elementary school in Brazil.

b) Text segments: the software divided the text into 129 text segments.

c) Number of active and supplementary forms: Words considered active (adjectives, nouns, verbs and adverbs) and supplementary (articles and pronouns). Articles and prepositions were eliminated.

d) Number of occurrences: Total number of words contained in the corpus.

e) Number of lemmas: differs from the number of forms, because the lemmas are the lemmatized forms, that is, the process, effectively, of deflecting a word to determine its lemma (the inflections are lexemes).

f) Average forms per segment: Number of occurrences divided by number of texts.

g) Number of classified segments: In the present case, 81.32% of the segments were classified due to the choice of word categories in the preferences menu (first menu presented in this analysis), as well as the choice of way of selecting text segments.

h) Number of clusters: number of classes determined by the analysis.

Importantly, to be useful for classifying any textual material, DHC type analyses require a minimum retention of 75% of the text segments. When an analysis is lower than this value, it is not considered an adequate analysis, as it offers only a partial classification (Camargo; Justo, 2013). In this sense, the textual corpus used for the analysis of the present study is considered representative and useful, as the use was 81.32%.

The interpretation of the DHC results is based on the hypothesis that the use of similar lexical forms is linked to common representations or concepts (Reinert, 1987). Thus, in the DHC tab of the IRaMuTeQ results, it is possible to access the phylogram (Figure 3) which presents the partitions that were made in the corpus until the final five classes were reached, as well as the words that configure each one. The phylogram is read from left to right.

In the classification results obtained using the Reinert Method (Phylogram, Figure 3), the corpus, or 'body,' was initially divided (first partition or iteration) into two subcorpus. This separation isolated Class 2, which represents 19.6% of the textual corpus. In the second partition, the subcorpus was divided as follows: 1) Classes 3 and 5 that indicate, respectively, 19.6% and 21.7% of the total; and 2) Classes 1 and 4 that indicate, respectively, 20.6% and 18.5% of the total.

We reinforce that the five classes contain the active forms or organized words that presented the highest frequency, in descending order, and that were significant to represent each of the subcorpus through the chi-square association test generated in the IRaMuTeQ reports. That is, the greatest adherence of them in the class and

between classes and that can be observed in the phylogram, which is another one of the DHC results by Reinert's Method (Figure 4).

Class 2 can thus be described as using state-of-the-art bibliographic studies (bibliographic study) and textbook analysis. Aspects of Probabilistic literacy according to Gal (2005) are indicated when he proposed a model composed of cognitive and dispositional elements (student attitudes towards knowledge: criticality, beliefs and attitudes and personal feelings). In addition, cognitive elements formed by big ideas (variation, randomness, independence, predictability and uncertainty), probabilistic calculations, language, context and critical issues were addressed.

Figure 3 - Result of Classification by Reinert's Method: Phylogram
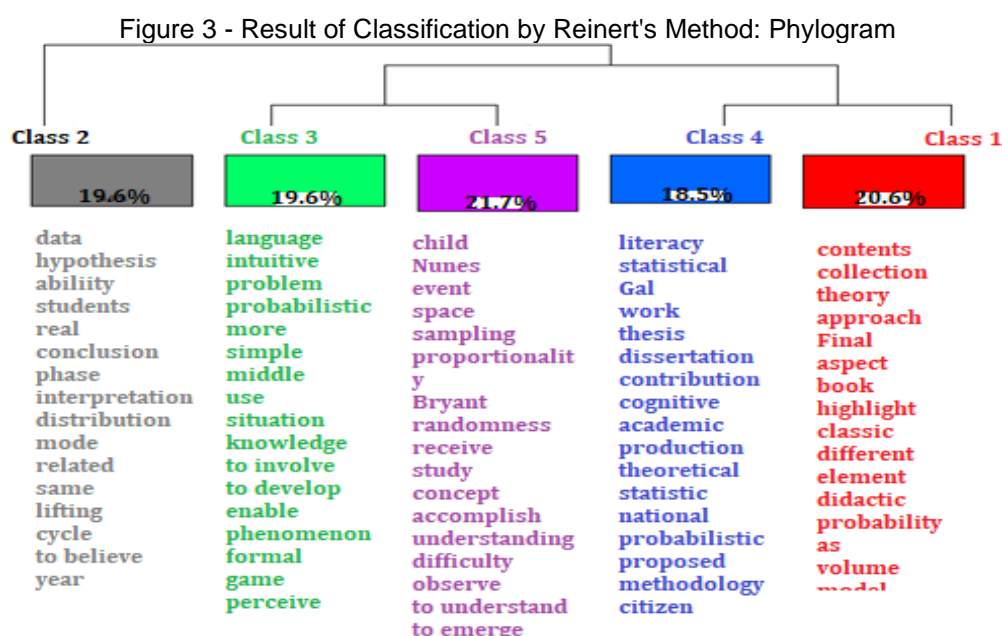


Source: IraMuTeQ software output, 2023

In this grouping, we have Assis (2018), text 4, which evaluated a collection of Elementary School books, based on four mathematics books, in which it was observed that a limited approach to elements of Probability Theory is proposed, using only classic contexts, which may lead to the emergence of learning difficulties, hindering the development of students' probabilistic thinking. As a positive point, the fact that the collection of mathematics textbooks addressed probability content in all volumes was highlighted, although a formal probabilistic language was observed, which is considered to be a hindrance to progressive learning and an obstacle to development of students' probabilistic thinking.

For this reason, Assis (2018) suggests that teachers implement activities that involve the use of roulette wheels, which are tools that enable a much broader approach to teaching probability through the study of non-equiprobable probabilistic phenomena, as well as how to contextualize probabilistic problems in a perspective closer to reality, making use of quotidian situations.

The text by Paim (2019), text 5, is also relevant. It presented a state-of-the-art review, exposing a descriptive and qualitative analysis of data related to statistical and/or probabilistic literacy present in Brazilian theses and dissertations, from 2013 to 2018, developed in contexts of elementary school, high school, initial and/or continuing teacher training in mathematics education. It is concluded that probabilistic and/or statistical literacy is not yet being achieved, but it is highlighted that there is an understanding that probabilistic language needs to be inserted through tasks that use natural language, as it demonstrates mistakes in the use of probabilistic language when emitting different degrees of uncertainty.

In Class 3 and 5, according to the grouping of results generated by DHC, the use of probabilistic knowledge of children in situations of games and probabilistic activities with everyday situations stands out, bringing to light a natural language, based on beliefs and opinions.

The active forms with the highest chi-square values in this class reveal the focus on concepts and language associated with the notion of randomness, random and deterministic phenomena, sample space, random events and the intuitive conception of chance. It also indicates that children use intuitive language to engage with the presented concepts. It is supported by Bryant and Nunes (2012) as well by the BNCC in Brasil (2018) and the National Curriculum Parameters - PCN (Brasil, 1997).

Thus, in Amaral (2007), text 1, the assumptions of Didactic Engineering were used as a methodology in which eleven students from the fourth and sixth semesters of the Pedagogy course at a private college in Jacareí (São Paulo, Brazil) participated in the activities. Data were analyzed in light of theoretical references on the formation of statistical thinking, levels of statistical and probabilistic literacy and statistical reasoning. It is concluded that the levels of probabilistic literacy at the cultural and functional levels do not necessarily align, highlighting the need for initial and continued training work for teachers in the early years of elementary school.

Amaral (2007) suggests the following procedures to support the construction of meanings for statistical and probabilistic concepts for students of Pedagogy courses: 1) promoting a teaching and learning process that begins with children's intuition, such as language, considering, therefore, the knowledge already acquired, their opinions and the relationships they are able to establish; 2) generate teaching activities that can offer their students the opportunity to carry out random experiments and associate their properties, establishing relationships to systematize different concepts.

In text 2, Silva (2016) studied the intuitive meaning of probability as evidenced by 36 children in the 1st, 3rd and 5th grade of elementary school. This effort brought to light a natural language, relating randomness to luck or chance, justifying the answers based on particular parameters and demonstrated better understanding in unlikely and impossible events.

Supported by Bryant and Nunes (2012), it is indicated that, although children easily eliminate impossible events, they often struggle to discriminate and compare highly improbable events with completely impossible ones.

It is also highlighted, based on Gal (2004), that students must understand the "language of chance", that is, the various ways of representing and communicating possibilities and probabilities The expressions and words used in everyday life may have an implicit semantic load that entirely differs from that proposed at the heart of a probabilistic question. Students should have the opportunity to describe their thinking and understanding of probabilistic concepts both orally and in writing.

The intuitive meaning of probability, as indicated by Batanero and Diaz (2007), was also evidenced by the children, especially because the procedure adopted was the manipulation of chance generators, in this case, dice and coins in game situations. This process revealed a natural language rooted in beliefs and opinions.

In addition to these two texts, Kian (2021), text 7, is also noteworthy for conducting multivariate textual analysis using the IraMuTeQ software. This analysis identified and explained the multiplicity of terms, oral and written expressions, symbols and representations used if you want students to learn the concept gradually and acquire the respective basic concepts in probability. The results revealed a predominance of common language words and verbal expressions, mainly related to the intuitive meaning of probability. Moreover, with the resolution of the sequence of

activities, Kian noted that students begin to appropriate a less intuitive language and move towards the use of a more formal language of probabilistic terms, necessary for mastering probabilistic concepts.

Finally, we present the grouping formed by Classes 1 and 4, which we call the use of skills related to hypotheses formulation, interpretation of real univariate and bivariate data, confrontation between hypothesis and data, evaluation of conclusions, and the use of probabilistic language in predictions. These classes are strongly linked with Class 5, as they present another aspect of the study by Bryant and Nunes (2012), alerting to the demands of correlational reasoning, which arise from the understanding of probability and proportionality.

This grouping is formed by the works of Cavalcanti and Guimarães (2018) and Cavalcanti (2019), texts 3 and 6. These studies examined elementary school students' understanding when raising hypotheses, analyzing real data and making decisions. They also explored the learning of students of this same teaching cycle on raising hypotheses, data analysis and conclusions from statistical data.

Bringing more detail, in Cavalcanti and Guimarães (2018), text 3, the understandings demonstrated by elementary school students (5th and 7th years) were analyzed in the face of univariate and bivariate distributions. When raising hypotheses, it was noticed in the writings of these students that the use of probabilistic language generated difficulties in situations of uncertainty, affecting the analysis of the expression of ideas. The conclusions indicate that difficulties in the use of probabilistic language demonstrate the need for pedagogical interventions aimed at analyzing and re-analyzing data to contribute to advances in the understanding of these skills.

In Cavalcanti (2019), text 6, the learning presented by elementary school students related to the hypothesis, data analysis and skills related to the conclusions by 5th and 7th grade students at Brazilian public school based on evidence related to the probabilistic language identifying difficulties of both years in all activities.

The results obtained with the Reinert method (DHC) by the Confirmatory Factor Analysis (CFA) also revealed relationships between the context variables and the classes resulting from the DHC (Figure 4). Remember that the variables used were: publication type (pubType); year of publication of the text (yearPub); public to which

the study is directed (publicTeaching); work focus (workFocus); and Brazilian region where the study was developed (PubRegion).

Figure 4 - Result of Classification by Reinert's Method: CFA, associated with the highlighted variables
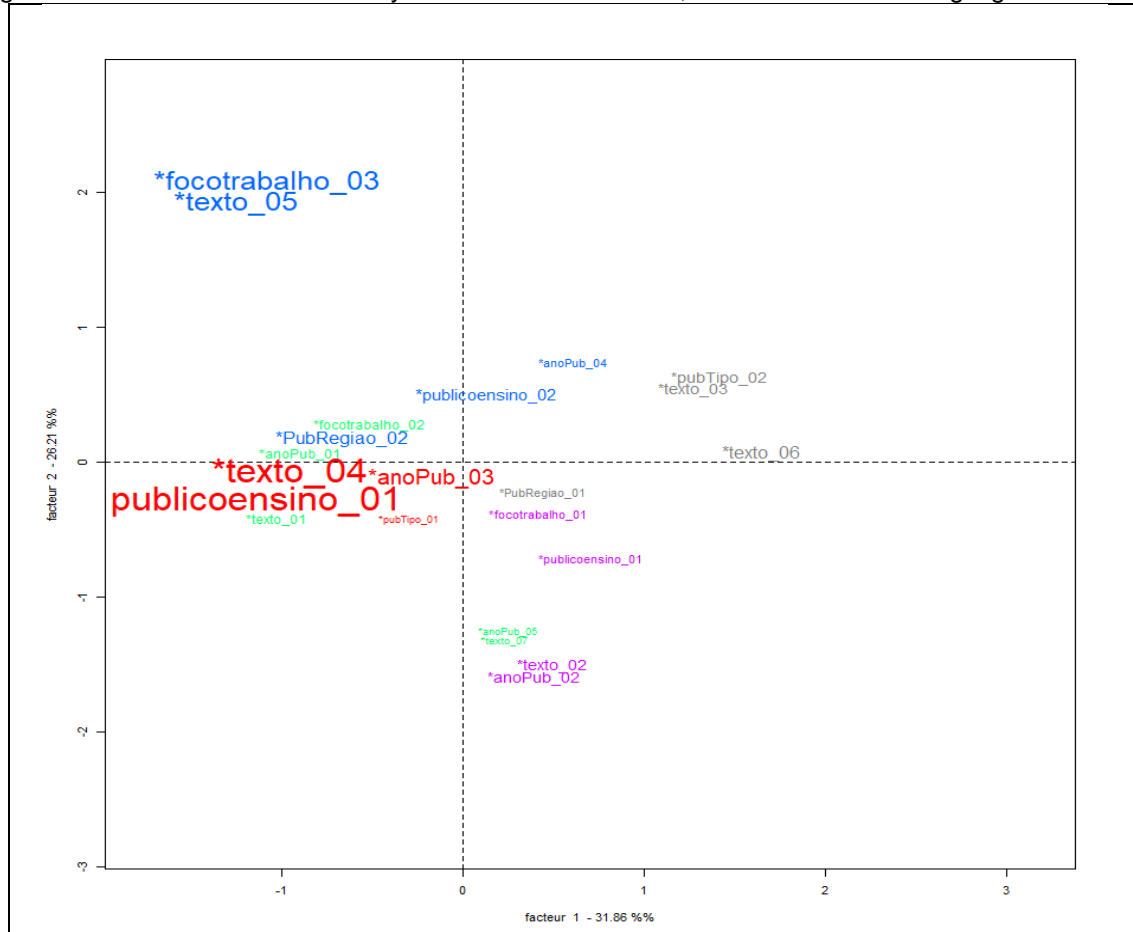


Source: IraMuTeQ software output, 2023

In what follows, considering Figure 4, we will present in detail which variables were used to identify each of the texts that are significantly associated with each of the classes generated by DHC.

In Class 2, "Use of bibliographic studies as state of the art and analysis of textbooks", formed by texts 4 and 5, respectively by Assis (2018) and Paim (2019), the focus was on document analysis. It is identified that: 1. It was carried out in the Northeast and Southeast regions of Brazil; 2. It is aimed at elementary school students and teachers in initial and continuing education; 3. Production published in Master's dissertation; 4. Focused on teaching probability or combining elements of statistics and probability.

When we consider Classes 3 and 5, referring to "Use of probabilistic knowledge of children in situations of games and probabilistic activities with everyday situations, bringing to light a natural language, based on beliefs and opinions", it is formed by texts 1, 2 and 7, respectively by Amaral (2007), Silva (2016) and Kian (2021). These studies were focused on the evaluation of probabilistic language through games and the presentation of random phenomena, identifying that: 1. It was performed in the Southeast and Northeast regions of Brazil; 2. Aimed at students in the early years of Elementary School and teachers in initial training (Pedagogy); 3. Production published in Master's dissertation; 4. Focused on teaching probability or combining elements of statistics and probability.

Finally, referring to Classes 1 and 4, referring to "Use of skills related to raising hypotheses, interpreting real univariate and bivariate data, confronting hypothesis and data, evaluating conclusions and using probabilistic language in predictions ", formed by the texts by Cavalcanti and Guimarães (2018) and Cavalcanti (2019), texts 3 and 6, the scope relied on the use of probabilistic language in predictions and it is identified that: 1. It was carried out in the Northeast region of Brazil; 2. Aimed at students in the early years of elementary school; 3. Based on a Doctoral thesis and scientific article resulting from its development; 4. Specifically focused on teaching Probability.

## 5 FINAL CONSIDERATIONS

In this text, we presented an analysis of scientific productions in the Brazilian context on the use of approaches to assess probabilistic language in the early years of elementary school. The inventoried works offer important elements for understanding the production of knowledge and highlight the contributions and gaps in this area of research, while also enabling this field to be further explored by researchers.

During the period from 2007 to 2021, seven studies were found that focused on the use of approaches to assess probabilistic language in the early years of elementary school. We seek to report a Systematic Literature Review (SLR) aimed at providing an overview of research development in relation to this theme in Brazil.

The set of works analyzed reveal important aspects:

1. The surveys conducted in the period 2007-2021 are concentrated in the Southeast and Northeast regions, with no investigation in the other Brazilian regions (South, North and Midwest).

2. Most research is aimed at students in the early years of elementary school and in initial teacher training (Pedagogy), lacking work more aimed at teachers in effective exercise of their profession.

3. Research has been primarily focused on teaching probability or combining elements of statistics and probability.

4. The vast majority of works were developed in postgraduate programs, including four Dissertations and a Doctoral thesis and a scientific article resulting from partial results of the thesis.

5. We emphasize that research on the use of probabilistic language assessment approaches in the early years of elementary school is still incipient, considering that the teaching of probability was inserted into the Brazilian curriculum structure, mainly in Basic Education with PCN, from 1997 (1st and 2nd cycles) and reinforced in the BNCC in Brazil (2018), when it is indicated that from the age of 6, aspects related to the notion of bad luck should be addressed.

6. In the maximum tree generated in IRaMuTeQ on the set of works, the concern in research regarding the student's profile and their relationships with experiential knowledge and everyday life is evident, contributing to the formation of citizens aware of the development of probability.

We indicate that, although research in Brazil is still incipient concerning the use of probabilistic language assessment approaches in the early years of elementary school, the works selected for this study show that Probability can influence the opinion formation of citizens, and it is essential that knowledge be built through the information available.

However, there is a need for further research that addresses the use of probabilistic language assessment approaches in the early years of elementary school in the Brazilian context.

Finally, the works recognize that probabilistic information is always available to society through the means of communication and, based on this assumption, it is considered that students already have some knowledge about the subject and, therefore, investigating the evaluation of the appropriation of a probabilistic language can help students in the construction of a new critical and autonomous knowledge through the availability of new knowledge.

**AILTON PAULO DE OLIVEIRA JÚNIOR**

Doutorado e Pós-Doutorado em Educação pela Universidade de São Paulo e Mestrado em Pesquisa Operacional pelo Instituto Militar de Engenharia. Atualmente é professor Associado IV do curso de Licenciatura em Matemática e do Programa de Pós-Graduação (Mestrado e Doutorado) em Ensino e História das Ciências e da Matemática da Universidade Federal do ABC - UFABC.

**FÁTIMA APARECIDA KIAN**

Doutoranda e Mestra no Programa de Ensino e História das Ciências e da Matemática da Universidade Federal do ABC – UFABC. Graduação em Direito e Pedagogia pela Universidade Nove de Julho (UNINOVE), História pela Universidade Metropolitana de Santos (UNIMES) e LETRAS pela Faculdade de Filosofia e Letras da Fundação Santo André.

**LUZIA ROSELI DA SILVA SANTOS**

Doutoranda e Mestra no Programa de Ensino e História das Ciências e da Matemática da Universidade Federal do ABC – UFABC. Graduação em Ciências de dados pela UNIVESP, Licenciatura em Pedagogia pela UNIJALES, Licenciatura em Artes Visuais pela UNIMES. Atua como Professora de Atendimento Educacional Especializado na educação fundamental I e II.

**REFERENCES**

AMARAL, M. H. (2007). *A estatística e a formação inicial com alunos de um curso de pedagogia*: reflexões sobre uma sequência didática. 2007. Dissertação (Mestrado em Educação) - Pontifícia Universidade Católica de São Paulo, São Paulo, Brasil, 2007.

ASSIS, J. de L. *Ensino de probabilidade:* análise de uma proposta para os anos finais do ensino fundamental. 2018. Dissertação (Mestrado em Ensino de Ciências e Matemática) - Programa de Pós-Graduação em Ensino de Ciências e Matemática da

Universidade Estadual da Paraíba – UEPB, Campinas Grande, Paraíba, Brasil, 2018.

BATANERO, C.; DIAZ, C. Meaning and understanding of mathematics. The case probability. *In*: FRANÇOIS, K.; VAN BENDENGEN, J. P. (eds.). *Philosophical Dimensions in Mathematics Education*. New York: Springer, 2007. p. 107-128.

BRASIL. *Base Nacional Comum Curricular*: Educação é a base. Brasília. Available in:
http://basenacionalcomum.mec.gov.br/images/BNCC_EI_EF_110518_versaofinal_site.pdf. Access in: 10 jan. 2023.

BRASIL. Secretaria de Educação Fundamental. *Parâmetros curriculares nacionais:* Matemática, v. 3 (Ensino Fundamental). Brasília: MEC, 1997.

BRYANT, P.; NUNES, T. *Children's understanding of probability*: a literature review. London: Nuffield Foundation. 2012. Available in:
http://www.nuffieldfoundation.org/sites/default/files/files/Nuffield_CuP_FULL_REPORT v_FINAL.pdf. Access in: 12 jan. 2023.

CAMARGO, B. V.; JUSTO, A. M. IRaMuTeQ: Um software gratuito para análise de dados textuais. *Temas em Psicologia*, Ribeirão Preto, SP, v. 21, n. 2, p. 513-518, 2013.

CAVALCANTI, E. M. S. *Aprendizagem de estudantes do ensino fundamental sobre levantamento de hipóteses, análise de dados e conclusões a partir de dados estatísticos*. 2019. Tese de Doutorado em Educação Matemática e Tecnológica, Universidade Federal de Pernambuco – Centro de Educação, Recife, Brasil, 2019.

CAVALCANTI, E. M. S.; GUIMARÃES, G. Compreensões demonstradas por estudantes do ensino fundamental ao levantarem hipóteses, analisarem dados reais e tomadas de decisões. *ReBECEM*, Cascavel, PR, v. 2, n. 2, p. 194-216, 2018.

FINDLEY, K.; WUITACRE, I.; ATABAS, S. What's in a name? Leveraging students' informal vocabulary in probability. *Teaching Statistics*, v. 43, n. 7, p. 1-10, 2020.

GAL, I. Towards 'probability literacy' for all citizens. *In*: JONES, Graham (ed.). *Exploring probability in school*: Challenges for teaching and learning. Kluwer Academic Publishers, 2004. p. 43-71.

GAL, I. Towards probability literacy for all citizens: Building blocks and instructional dilemmas. *In*: JONES, G. (ed.). *Exploring probability in school*. US: Springer, 2005. p. 39-63.

GÓES, M. C. R.; CRUZ, M. N. Sentido, significado e conceito: notas sobre as contribuições de Lev Vygotsky. *Pro-Posições*, Campinas, SP, v. 17, n. 2, p. 31-46, 2006.

JUSTO, A. M.; CAMARGO, B. V. Estudos qualitativos e o uso de *softwares* para análises lexicais. *In*: NOVIKOFF, Cristina; SANTOS, S. R. M.; MITHIDIERI, *Otávio Barreiros* (orgs.). *Caderno de artigos:* X SIAT & II Serpro. Rio de Janeiro: Lageres, 2014. p. 37-54.

KIAN, F. A. *A linguagem probabilística no final dos anos iniciais do ensino fundamental: um caminho para o desenvolvimento da alfabetização probabilística*. 2021. Dissertação (Mestrado em Ensino e História das Ciências e da Matemática) – Universidade Federal do ABC, Pós-graduação em Ensino e História das Ciências e da Matemática, Santo André, São Paulo, 2021.

KITCHENHAM, B. et al. *Guidelines for performing systematic literature reviews in software engineering*. Technical Report EBSE 2007-001, Keele University and Durham University Joint Report. 2007.

LEBLANC, J.-M. Proposition de protocole pour l'analyse des données textuelles: Pour une démarche expérimentale en lexicométrie. *Nouvelles perspectives en sciences sociales* (NPSS), Grenoble, v. 11, n. 1, p. 25–63, 2015.

MUTOMBO, E. *A bird's-eye view on the EC environmental policy framing*. 10 years of Impact assessment at the commission: The Case of DG ENV: ICPP 2013. 1st International Conference on Public Policy, Grenoble, 2013. p. 26-28.

NASCIMENTO, A. R. A.*;* MENANDRO, P. R. M. Análise lexical e análise de conteúdo: uma proposta de utilização conjugada. *Estudos e Pesquisas em Psicologia*, Rio de Janeiro, RJ, v. 6, n. 2, 72-88, 2006.

OLIVEIRA, L. F. R. *Tutorial (básico) de utilização do IRaMuTeQ*. [Goiânia]: Universidade Federal de Goiás, Brasil, 2015. Available in: https://files.cercomp.ufg.br/weby/up/771/o/Tutorial_-_Revis%C3%A3o.pdf. Access in: 11 jan. 2023.

PAIM, S. A. de O. C. *O estado da arte das pesquisas brasileiras sobre o letramento estatístico e probabilístico*. 2019. Dissertação (Mestrado em Ciências Exatas) - Universidade Federal de São Carlos – UFSCar, Centro de Ciências Exatas e Tecnologia, São Carlos. São Paulo, Brasil, 2019.

REINERT, M. *Alceste*: Analyse de données textuelles. Manuel d'utilisateur. Toulouse: IMAGE, 1998.

REINERT, M. Classification descendante hiérarchique et analyse lexicale par contexte: application au corpus des poésies d'Arthur Rimbaud. *Bulletin de méthodologie sociologique*, v. 13, n. 1, 1987.

SALVADOR, P. T. C. de O. et al. IRaMuTeQ nas pesquisas qualitativas brasileiras da área da saúde: scoping review. *Revista Brasileira em Promoção da Saúde*,

Fortaleza, CE, v. 31, p 1-9, 2018. Available in:
https://ojs.unifor.br/RBPS/article/view/8645/pdf. Access in: 11 jan. 2023.

SILVA, R. de C. B. *É a moeda que fiz, não é a gene que quer não:* conhecimentos probabilísticos de crianças em situações de jogos. 2016. Dissertação (Mestrado em Educação Matemática e Tecnológica) - Universidade Federal de Pernambuco, Centro de Educação, Recife, Brasil, 2016.

TORRES, E. G.; HARO, J. J. O.; BATANERO, C.; CONTRERAS, J. M. El lenguaje de Probabilidad en los libros de texto de educación primaria. *Revista Iberoamericana de Educação Matemática*, España, v. 35, p. 75-91, 2013.

VAN DOOREN, W.; DE BOCK, D.; JANSSENS, D.; VERSCHAFFEL, L. The illusion of linearity: expanding the evidence towards probabilistic reasoning. *Educational Studies in Mathematics*, v. 53, n. 2, p. 113–138, 2003.

VÁSQUEZ, C. O. *Evaluación de los conocimientos didáctico-matemáticos para la enseñanza de la probabilidad de los profesores de educación primaria em activo*. 2014. Tesis Doctoral en Educación, Programa de Doctorado em Educación, Universitat de Girona, España, 2014.

VÁSQUEZ, C. O. Surgimiento del lenguaje probabilístico en el aula de educación primaria. *REnCiMa*, São Paulo, SP, v. 9, n. 2, p. 374-389, 2018.

VYGOTSKY, L. S. *A construção do pensamento e da linguagem*. Tradução Paulo Bezerra. 2. ed. São Paulo: Martins Fontes, 2009.