EDUCATIONAL EVALUATION AND ARTIFICIAL INTELLIGENCE: THE VIEW OF FUTURE TEACHERS

Nicolas Valverde Costa[1] and Paulo Sérgio Garcia[2].

ABSTRACT

The aim of this study was to investigate the perceptions of future teachers of a Pedagogy course regarding the use of ChatGPT and its application in the context of educational evaluation. This is an exploratory research carried out at a public university located in the Greater ABC Paulista region (São Paulo), which sought to understand how students perceive the potential, challenges, and ethical and pedagogical implications of the use of this technology in evaluation processes. To this end, a mixed methodology was used, with the first qualitative phase, the use of interviews, and the second quantitative, the use of a Likert-type scale. The results revealed an ambivalent view on the use of ChatGPT in educational assessment. On the one hand, participants highlight its potential to support pedagogical practices, such as planning and diversification of instruments. On the other hand, items raise concerns about dehumanization and loss of teacher autonomy. This highlights the need for caution and critical use of the tool, preserving the centrality of the teacher. Responsible Chat integration must be accompanied by ethical and reflective training to maximize its benefits.

Keywords: Evaluation. Artificial intelligence. ChatGPT. Teacher Training.

[1] Student of Pedagogy and Scientific Initiation
Municipal University of São Caetano do Sul
[2] Dr. in Education from the Faculty of Education of the University of São Paulo
Full Professor of undergraduate and graduate courses at
Municipal University of São Caetano do Sul

INTRODUCTION

The assessment of learning, as it is currently found, especially in elementary, secondary and higher education, has predominantly standardized and homogeneous characteristics, following a format that tends to ignore the complexity of the educational and social contexts of students. These evaluation practices occur in a superficial and momentary way, focusing more on the punctual measurement of academic performance than on the student's formative procedure. Often, this process distances itself from social realities and needs, becoming more of a formal exercise and limited to traditional instruments, such as tests and written tests.

In this scenario described, with these particularities, authors indicate (Swiecki *et al.*, 2022) that Artificial Intelligence (AI), which has expanded into several areas of application simultaneously, covering fields such as medicine, law, business, education, among others, has the potential to approach and answer some of the questions raised above, even if, at this time, it does not represent definitive solutions.

In fact, AI has, according to researchers (Beaulac; Rosenthal, 2019), a great potential to promote more personalized learning, adaptive and automated assessments, and intelligent tutoring systems. In addition, it enables the integration of virtual and augmented reality elements in the educational environment, enriching the learning experience. It also contributes to data analysis with a focus on performance forecasting, language teaching, and promoting accessibility and inclusion, expanding the reach and effectiveness of educational processes. Its growth is associated with the ability of its algorithms to make recommendations, predictions, and make decisions, favoring learning and improvement processes in different contexts (Chen *et al.,* 2022).

The field of evaluation, based on Artificial Intelligence, has been explored in several international studies (Montagner, 2019; Beaulac; Rosenthal, 2019; Chen, Chen et al., 2020; Colônia *et al.*, 2020; Alam, 2021; González-Calatayud, et al., 2021; Swiecki *et al.*, 2022; Gardner; O'Leary; Yuan, 2021; Chen et al., 2022; Hooda *et al.*, 2022; Martínez-Comesana *et al.* 2023; among many others) and national (Campos; Lastória, 2020; Preuss *et al.* 2020; Lima *et al.*, 2020, among others), performing analyses and conjectures, in addition to showing the potential of AI in evaluation practices.

In the field of AI, *ChatGPT* is indicated as an element that will enable various opportunities in the field of education. For example, it can promote peer-to-peer collaboration and facilitate the sharing of educational resources, allowing teachers to

connect and collaborate in innovative ways (Marcom; Porto, 2023), and enable professional development, based on collaboration between educators and resources for their pedagogical practices (Duque *et al.*, 2023).

Specifically, in the field of educational evaluation, a recent survey (ObeducGABC, 2024) carried out with basic education teachers revealed that *ChatGPT* is considered a tool that can support teacher training with regard to the topic of evaluation. The professors who participated in the study identified the potential of the platform to offer practical guidance, clarify doubts and deepen the understanding of evaluation strategies, making it a resource for professional development. Other studies have followed the same direction (Lo, 2023).

The research conducted by ObeducGABC (2024) shows that the use of ChatGPT transcends the mere facilitation of access to content and practices related to the assessment of learning, consolidating itself as a relevant tool in strengthening the understanding and application of fundamental concepts, such as *feedback*. However, some issues should be treated with caution: it is necessary to develop evaluative judgment processes in teachers (Tai *et al.,* 2018).

Considering that the introduction of AI in education is recent, especially in the field of evaluation, this study aimed to investigate the perceptions of future teachers of a Pedagogy course in relation to the use of *ChatGPT* and its application in the context of educational evaluation.

EVALUATION AND ARTIFICIAL INTELLIGENCE: EXPLORING THEIR RELATIONSHIPS

The term "Artificial Intelligence" was coined by John McCarthy, a pioneer and one of the founders of the field, in 1956. It is an area of computer science focused on the development of systems or machines that can perform tasks that, in humans, would require intelligence. These tasks include, among others, pattern recognition, decision-making, natural language interpretation, and the ability to learn from data. The goal of AI is to simulate or replicate human reasoning and learning processes, using algorithms, neural networks, and other computational tools (Russell; Norvig, 2016).

For Russell and Norvig (2016), AI encompasses several areas of study and application, each with specific objectives and techniques. Among the most relevant we have machine learning (*Machine Learning*) -- which enables systems to learn from data to make predictions and decisions autonomously --, Natural Language Processing (NLP) -- aimed at enabling the understanding and interaction between computers and human

language, allowing machines to understand, interpret and generate text or speech, facilitating more fluid communication with users. In addition, as further developed in studies by Torfi *et al.* (2020) and Otter *et al.* (2021), NLP also covers the translation and manipulation of texts.

The authors also highlight "Computer Vision" – an area focused on the ability of machines to extract and interpret visual information from images or videos, replicating the human capacity for visual perception, and Robotics – an area that seeks to create intelligent systems capable of performing physical activities in the real world, integrating aspects of control and autonomy in varied environments. We also have the so-called "Symbolic AI" and Expert Systems, which address the representation of knowledge and decision-making based on rules and structured logic, as well as Artificial Neural Networks (ANN) and *Deep Learning*, fields that represent an evolution of these capabilities, simulating human brain processes to develop deeper and more complex learning models.

Regarding the impact of AI on education, we can divide it into three main areas, according to Martínez-Comesana (2023): administration, instruction, and learning; in the latter, elements that add value to AI stand out, especially through machine learning, based on statistical modeling and cognitive theory (Chen, Chen *et al.*, 2020).

Other authors (González-Calatayud, Prendes-Espinosa, and Roig-Vila, 2021), in turn, point out that the application of AI in the educational field has been concentrated, above all, in the areas of tutoring and evaluation. It has also been used to generate automatic grades, classify students, and provide different types of *feedback*, contributing to a more adaptive and effective educational approach. In this sense, Hooda et al. (2022) investigated the impact of assessment and *feedback* on academic results and student performance in higher education, highlighting the role of AI in improving the learning experience and developing students' skills. Similarly, Martínez-Comesana *et al.* (2023), in their analysis, highlight that the main contributions of the application of AI in student assessment in basic education include the ability to predict academic performance, provide automated *feedback*, and make assessment processes more objective and meaningful.

Alam (2021) demonstrated that, currently, up to 40% of teaching time is dedicated to activities that can be automated, indicating a considerable potential for expansion for the use of AI in this sector. In this context, AI has the potential to "take over" a part of teachers' tasks related to the assessment of learning, especially in repetitive processes and data analysis. AI can automate the grading of tests and assignments, provide instant *feedback*

to students, and tailor assessments according to each student's individual progress, which allows for greater personalization of teaching. In other words, AI tools can assist in identifying patterns in student performance, facilitating faster and more effective pedagogical intervention, and thus, by reducing the operational burden on teachers, they would have more time to focus on qualitative aspects of teaching, such as skills development and personalized support (ObeducGABC, 2024).

Another highlight of AI is the electronic platforms for online and offline exams, offering multimedia questions that would not be feasible in traditional assessments (with the exclusive use of paper) and organizing them in a predetermined or random way, with quick and personalized feedback for students. These adaptive platforms create personalized learning paths based on students' prior knowledge, promoting cognitive development and autonomy.

Another possibility is related to the development of authentic assessments, which use tasks designed to simulate real situations. AI techniques are being implemented to enrich these simulations and assess evidence of student performance. In both virtual and face-to-face learning contexts, AI has been playing a central role, being essential for the creation and evaluation of these realistic and applicable experiences in the real world (Swiecki *et al.,* 2022).

However, investigations on the application of AI in the field of educational evaluation remain at an early stage in Brazil. Fernandes *et al.* (2024) analyzed the impacts of AI-mediated processes on the academic performance of high school students, highlighting promising possibilities for personalization and optimization of both pedagogical and evaluative practices, reinforcing the findings of the aforementioned studies. In a complementary perspective, Campos and Lastória (2020) emphasized the importance of AI in the personalization of learning experiences, exploring the use of algorithms aimed at teacher evaluation and the adoption of gamification strategies as catalysts for the educational process. Also in the same vein, Preuss et al. (2020, p. 443) highlighted the use of artificial intelligence in interactive platforms aimed at inclusive education, highlighting the potential of this technology to promote more accessible and adaptive assessments, aligned with the various educational demands. In addition, Lima *et al.* (2020, p. 68) highlight that the integration of serious games with AI-based evaluative approaches not only contributes to raising student motivation, but also deepens the understanding of concepts related to artificial intelligence, enhancing engagement and meaningful learning in the school context.

As a summary, it can be stated that AI has the potential to "revolutionize" the field of learning assessment by overcoming initial challenges related to its implementation and by being understood and used efficiently by teachers. Through adaptive assessment tools, AI adjusts activities to the pace and needs of each student, promoting personalization, inclusion and dynamism. Additionally, AI makes it easier to create authentic simulations, increasing engagement and making assessments more inclusive for students with special needs. With immediate and detailed feedback, AI enables effective pedagogical interventions and helps students identify and overcome difficulties, transforming assessment into an adaptive process oriented towards the integral and continuous development of the student.

CHATGPT IN THE CONTEXT OF ARTIFICIAL INTELLIGENCE

Lo (2023) highlighted the potential of *ChatGPT* to contribute to pedagogical processes in various areas of education. In addition, studies by the Greater ABC Education Observatory (ObeducGABC, 2024) have highlighted promising applications of the tool in early childhood education, such as supporting pedagogical planning, developing playful activities aligned with the curriculum, offering continuous *feedback* on children's progress, creating personalized teaching materials, stimulating the imagination through adapted narratives, and promoting personalized teaching. In addition, the tool can facilitate communication with families and assist in the continuing education of teachers, suggesting updated and relevant pedagogical resources.

In elementary school, *ChatGPT* has proven to be a valuable resource in the preparation of organized lesson plans, with clear pedagogical objectives and activities that foster student engagement. The tool also supports the correction of textual productions, offering detailed feedback on aspects such as coherence, cohesion, grammar and structure, promoting the improvement of students' writing skills. It can also suggest interdisciplinary projects, integrating different areas of knowledge and contributes with support materials for review, reinforcing students' preparation for assessments and consolidating essential content (ObeducGABC, 2022).

Wang et al. (2021) emphasize the role of *ChatGPT* in foreign language teaching, especially through real-time interactions that simulate authentic conversations and assist in correcting errors in written production, favoring the development of students' language skills.

Also noteworthy is the potential of *ChatGPT* in creating interactive and personalized classes. In this context, the system acts as a "virtual tutor", identifying students' strengths and limitations and promoting a learning environment that is more adapted to individual needs. In this scenario, the proposal to integrate *ChatGPT-4* with the main *EdTech platforms*, such as *Khan Academy, Quizlet, and Duolingo*, stands out as a strategy to improve both the learning experience and the evaluation processes. With its ability to act as a personal tutor, teacher assistant, and automated *feedback* system, *ChatGPT* has significantly expanded the possibilities for personalized and interactive learning.

In the case of Khan Academy, for example, an experimental AI tool called Khanmigo is under development. The central objective is to offer students a virtual tutor capable of interacting through thought-provoking questions and meaningful conversations, encouraging active participation and critical thinking, while encouraging problem-solving and an in-depth understanding of the contents. In addition, these innovations also seek to alleviate the teaching workload, especially by automating evaluation practices and providing *real-time* feedback.

However, it is important to recognize that the incorporation of *ChatGPT* into education is still at an early stage, presenting challenges to overcome. The tool has limitations such as the difficulty in understanding nuances and specific contexts, which can compromise the accuracy and quality of the answers generated. Such challenges highlight the need for careful implementation and continuous studies to maximize the benefits of this technology in the educational field.

METHODOLOGY

The main objective of this study was to investigate the perceptions of future teachers of a Pedagogy course in relation to the use of *ChatGPT* and its application in the context of educational evaluation. This is an exploratory research carried out at a public university located in the Greater ABC Paulista region (São Paulo), seeking to understand how students perceive the potential, challenges, and ethical and pedagogical implications of the use of this technology in evaluation processes.

The choice for the exploratory character is due to the need to deepen the analysis on an emerging theme in the field of education, especially with regard to the role of artificial intelligence tools in the development of innovative evaluation practices. Thus, the present study adopted a mixed methodological approach, integrating quantitative and qualitative

methods in a complementary way. According to Cresswell (2007), the use of multiple data sources allows for a more comprehensive and in-depth understanding of the investigated phenomenon, employing both open and closed questions for data collection and analysis, including statistical and textual analysis procedures.

The study was structured in two stages. The first phase adopted a qualitative approach, characterized by an exploratory and interpretative approach, which uses the object of study as a starting point for the apprehension of new knowledge and the development of empirical theories. From this perspective, the valorization of the subjective point of view of the participants is prioritized, in line with the theoretical assumptions of Flick (2004). By opting for this methodology, the objective is not limited to the description of the phenomenon, but includes the interpretation of meanings, contexts and social relations, expanding the understanding of the elements under analysis in a denser and more integrated way.

In this phase, we interviewed 16 future teachers who were randomly selected, all enrolled in a Pedagogy course, three from the first, second and third semesters, two from the fourth and five from the eighth. The interviews collected data on the participants' profile and their perceptions of *CHATgpt* in relation to the evaluation.

The interviews were concluded based on the saturation index criterion, as conceptualized by Bertaux (1980). This criterion refers to the moment when, after conducting a sufficient number of interviews, the researcher achieves a comprehensive and consistent understanding of the object of study. At this stage, it becomes evident that the main similarities and differences between the collected data were explored and understood in a satisfactory manner, indicating that the continuity of data collection would not add substantially new information to the analysis.

The interviews were subjected to an analysis based on the theoretical guidelines of Bardin (2015, p. 40). Among the various techniques available for content analysis, the categorical one stands out, which is historically the oldest and most widely used. In the present study, we opted for evaluation analysis, whose purpose, according to Bardin (2015, p. 201), is "[...] measure the speaker's attitudes towards the objects he discusses". This approach allows interpreting the attitudes and perceptions of the interviewees, contributing to the understanding of the dynamics and meanings associated with the investigated phenomenon.

In the second stage, based on the data from the first, we developed a *Likert-type* scale with 35 items, ranging from 1 to 4 points, as follows: 1, I do not agree; 2, I agree a little; 3, I agree; 4, I strongly agree. From the elaboration of the items, five domains were created qualitatively by the researchers, namely: FSP (Teacher Support Tools): [2, 6, 8, 10, 16, 17, 25]; INA (Negative Impacts on Students): [3, 4, 13, 20, 28, 31, 32]; NPI (Negative Impacts on Teachers): [18, 19, 21, 33, 35]; IPA (Positive Impacts on Students): [5, 7, 11, 22, 24, 26, 29, 34]; IPTP (Positive Impacts on Teachers' Work): [1, 9, 12, 14, 15, 23, 27, 30].

In this phase, in which students from the 1st, 2nd, 3rd, 4th, 6th and 8th semesters participated, despite more than 150 students being invited, only 91 answered the questionnaire. In addition to the items, information on gender, age, and type of work was collected.

For the treatment of quantitative data, the application of descriptive statistics was adopted, using Microsoft Excel, an integral part of the *Office package*. The analysis followed the precepts described by Vieira (1997), including the calculation of measures of central position and dispersion. As a measure of central tendency, the arithmetic mean (μ) was used. To measure the variability of the data, the standard deviation (SD) and the coefficient of variation (CV) were calculated.

The coefficient of variation was interpreted based on criteria of high variability, considering values equal to or greater than 30% (CV = SD/μ). This measure allowed us to identify significant divergences among teachers in initial training, related to the items analyzed. In the present study, the analysis focused on the direction of the observed variations, in line with the objective of describing the phenomenon and interpreting the meanings attributed to the object of study.

The results were presented in tables, offering clearer visual representations. This analytical approach, combined with elements of the literature, enabled a more detailed discussion about the nuances and specificities of the challenges encountered by the students, promoting a deeper understanding of the phenomenon investigated.

RESULTS: PERCEPTIONS OF FUTURE TEACHERS REGARDING *CHATGPT* AND EVALUATION

The data from the interviews showed a profile of the future teacher, all of whom were women, with an average age of almost 22 years (SD=4.9) and workers in early childhood education (50%) and elementary education (50%)[3].

The data collected during the qualitative interviews played a central role in the construction of the Likert-type scale used in the study. These data allowed the identification of patterns, themes and relevant perceptions among the participants, providing essential support for the formulation of the scale items.

Based on this information, the domains that structure the scale were developed, as described in the methodological section. Each domain was designed to capture specific aspects of the object of study, ensuring that the items clearly and accurately reflected the most relevant dimensions identified during the analysis of the interviews. Table 1 summarizes the data from the scale:

Table 01: Profile of Participants

Category	Subcategory	%
Gender	Female	94.38%
	Male	4.49%
	Other	1.12%
Semester	2	25.84%
	6	19.10%
	4	19.10%
	3	15.73%
	1	14.61%
	8	5.62%
Working Level	Early Childhood Education	51.69%
	Elementary School	23.60%
	Other	11.24%
	Middle school	10.11%
	Does not work	3.37%

- Rounded data.

The profile of the study participants, with a mean age of approximately 24 years (SD = 8.1), reveals a predominance of women, reflecting the historical presence of women in pedagogical training. Most are distributed in the initial and intermediate semesters of the Pedagogy course, with less representation in the final periods. As for professional performance, it is observed that Early Childhood Education is the main field of work,

---

[3] Rounded data.

followed by Elementary Education and, to a lesser extent, by High School and other areas. Only a small portion of the participants declared that they were not working at the moment.

These data offer a comprehensive overview of the characteristics of future teachers, essential to understand the challenges faced in relation to the object of the study. Regarding teachers' perceptions of *ChatGPT* and assessment, Table 02 summarizes the results:

Table 02: Perceptions of future teachers regarding ChatGPT and assessment

Item	ChatGPT can...	Media	DP	CV	Gift.
8	... Provide fast, accurate information that can support teaching	3.09	0.92	0.3	FSP
9	... help teachers plan lessons more effectively	3.08	0.93	0.3	IPTP
15	... to enable teachers to explore more diverse assessment tools	3.06	0.97	0.32	IPTP
10	... improve creativity in the preparation of assessment tools	3.05	0.91	0.3	FSP
4	... facilitate plagiarism in student work	2.97	1.09	0.37	INA
6	... allow teachers to save time in the development of assessment tools	2.96	0.99	0.34	FSP
30	... help teachers design more inclusive and accessible assessment activities	2.93	0.96	0.33	IPTP
28	... increase the risk of standardized answers in tests, reducing the originality of students' answers	2.92	1.12	0.38	INA
1	... assist teachers in professional development	2.91	0.91	0.31	IPTP
3	... facilitate the students' 'cheating'	2.81	1.12	0.4	INA
12	... contribute to the development of more interactive and dynamic evaluation practices	2.8	0.95	0.34	IPTP
16	... reduce the time it takes to prepare assessment tools and quality criteria	2.77	1.04	0.38	FSP
27	... facilitate integration between different areas of knowledge in evaluation activities	2.76	0.99	0.36	IPTP
2	... be a tool that can assist teachers in the area of learning assessment	2.75	0.94	0.34	FSP
29	... Broaden the reach of assessments by allowing students to access reviews and feedback outside of the school setting	2.69	1.06	0.4	Wholesale price index
5	... in the assessment of learning, facilitate the personalization of teaching for each student	2.66	1.06	0.4	Wholesale price index
24	... promote the autonomy of students in the search for information and in the construction of knowledge	2.66	1.06	0.4	Wholesale price index
22	... encourage students' critical reflection on the use of technologies in education	2.62	1.03	0.39	Wholesale price index
17	... support in the correction of tests and activities	2.61	1.07	0.41	FSP
20	... making students overly reliant on technology in assessment	2.61	1.22	0.47	INA
26	... help personalize feedback for students of different proficiency levels	2.57	0.95	0.37	Wholesale price index
32	... encourage a superficial approach by students to learning, focusing only on quick responses	2.55	1.15	0.45	INA

13	... Lead students to read less	2.52	1.19	0.47	INA
25	... serve as a platform for test simulation and assessment practices	2.49	1.05	0.42	FSP
33	... create a distance for teachers from understanding individual student progress	2.44	1.18	0.49	INP
23	... assist teachers in identifying gaps in student learning	2.41	1	0.41	IPTP
7	... facilitate student engagement in assessments	2.3	1.05	0.46	Wholesale price index
11	... promote critical thinking among students by helping them explore new ideas	2.3	1.03	0.45	Wholesale price index
14	... change the learning assessment process	2.28	1.05	0.46	IPTP
18	... dehumanizing the evaluation process	2.18	1.15	0.53	INP
31	... reduce direct interaction between teachers and students in the evaluation process	2.16	1.1	0.51	INA
34	... encourage students to develop self-assessment skills	2.15	1.07	0.5	Wholesale price index
35	... replace the judgment made by teachers in the evaluation with automatically generated answers	2.11	1.06	0.5	INP
21	... limiting the autonomy of teachers in relation to the assessment of learning	1.9	1.06	0.56	INP
19	... replace the role of the teacher in the evaluation of learning	1.49	0.83	0.55	INP

The table shows that the highest-scoring items (average above 2.80) reflect the most positive perceptions of ChatGPT's potential in the context of the assessment. Among these items are: (8) Offer quick and accurate information that can support teaching: this is a support for teachers (average 3.09, whose domain is FSP); (9) Help teachers plan classes more efficiently: The tool is seen as a support to optimize lesson planning, evidencing its ability to assist in the teaching work (3.08, IPTP); (15) Allow teachers to explore more diversified assessment instruments: shows that several instruments can be developed (3.06, IPTP); (10) Improving creativity in the preparation of assessment tools (3.05, FSP); (4) Facilitate plagiarism in students' work: the possibility of the student taking negative advantage of the Chat is indicated (2.97, INA); (6) allow teachers to save time in the development of assessment instruments: support for teachers at work (2.96, FSP); (30) Helping teachers design more inclusive and accessible assessment activities: support (2.93, IPTP); (28) Increase the risk of standardized answers in tests, reducing the originality of students' answers: indicates that the tool can pose risks to students (2.92, INA); (1) to assist teachers in professional development: it brings the possibility of training (2.91, IPTP); (3) Facilitating the students' 'cheating': indicates the concern with learning (2.81, INA).

The best-rated items highlight both the positive potential and concerns related to their use in the field of evaluation. It is observed that the IPTP (Positive Impacts on Teachers' Work), FSP (Teacher Support Tools) and INA (Negative Impacts on Students) domains appear with almost similar frequency, representing, respectively, 40%, 30%, and 30% of the items analyzed. This suggests a balance between optimistic perceptions, focused on innovation and efficiency, and concerns about risks and limitations.

The predominance of the IPTP domain in the best-rated items reflects the perception that *ChatGPT* can offer innovative and practical contributions to evaluation, especially in planning, creativity, and diversification of instruments. However, the presence of items in the INA domain warns of risks related to academic integrity and the superficiality of the answers, indicating the need for strategies that mitigate such problems.

On the other hand, the table also brings the items with the lowest score (average below 2.20), which reflect concerns and negative perceptions about the impact of *ChatGPT* on the evaluation process. Among these items are: (18) Dehumanizing the evaluation process: Perceived as a risk, this item suggests that the use of *Chat* can make the evaluation less human, compromising its sensitivity and personalization (mean 2.18, INP domain); (31) Reduce direct interaction between teachers and students in the evaluation process: Raises concerns about the distance in pedagogical relationships caused by the use of the tool (2,16, INA); (34) Encourage students to develop self-assessment skills: Although it is a desirable goal, participants more critically evaluate the effectiveness of *ChatGPT* in promoting this competence (2,15, IPA); (35) Replacing the judgment made by teachers in the evaluation with automatically generated responses: Reflects concerns about the loss of human judgment in the evaluation process, seen as an essential element of teaching practice (2,11, INP); (21) Limiting teachers' autonomy in relation to learning assessment: This item suggests that the use of *ChatGPT* can restrict teachers' decision-making capacity, negatively affecting their autonomy (1.90, INP); (19) Replacing the role of the teacher in the evaluation of learning: It has the lowest average among all the items, indicating a strong rejection of the idea that technology can replace the teacher in this fundamental aspect of educational practice (1,49, INP).

These results, all in the domains of negative impacts, whether on teachers or students, point to a critical view of the participants in relation to the impacts of *ChatGPT* in the context of evaluation, especially with regard to the preservation of humanization, autonomy and centrality of the teacher's role in evaluation practices.

It should be noted that the averages of all the items analyzed, in general, ranged between 3.09 and 1.49, indicating considerable variability in the perceptions of future teachers about the use of *ChatGPT* in the context of evaluation. This breadth reflects both positive and optimistic perceptions and criticisms and concerns regarding technology.

The coefficient of variation, mostly higher than 0.30, indicates a high dispersion of the data in relation to the mean, demonstrating the heterogeneity of the responses among the participants. This high degree of variability can be interpreted as a reflection of significant divergences in individual perceptions.

These differences can also be explained by a number of factors, including variations in the degree of familiarity with the tool, the different ways of using it, previous experiences, as well as personal expectations and understandings about the potential benefits and risks associated with its use. Thus, the results reinforce the need to consider such disparities in the analysis and formulation of strategies to address the use of the tool in a more inclusive and effective way.

The data collected on the "positive and negative" elements, associated with the use of *Chat* in evaluation practices, provide a contribution to the panorama previously drawn by a study with few participants carried out by the study of the Education Observatory of the Greater ABC (ObeuducGABC, 2024). This previous study, conducted in a much more restricted context, pointed out the existence of differences in the understanding and appropriation of the tool within the scope of evaluation practices. The results of the present study, with many participants, by expanding the sample and exploring more varied nuances, not only corroborate the evidence raised by ObeducGABC, but also provide additional subsidies to understand the potentialities and limitations of *Chat* as an instrument to support evaluation. This broadening of perspective highlights the diversity of perceptions and experiences that permeate the use of the tool, reinforcing the need for more in-depth investigations that consider the multiple contexts and profiles of users.

FINAL CONSIDERATIONS

The data reveal an ambivalent view on the use of *ChatGPT* in the area of evaluation. On the one hand, the best-scored items show the recognition of the tool's potential to offer practical and innovative support to teachers, such as in lesson planning, diversification of assessment instruments and promotion of efficiency. These aspects,

predominantly in the IPTP and FSP domains, suggest that technology can be an important resource when used to improve evaluation practices.

However, the items with lower means reflect important concerns related to the risks of dehumanization, loss of teacher autonomy and negative impact on teacher-student interaction. These items, concentrated in the domains of INA and INP, reinforce the need for caution and critical reflection in the use of *ChatGPT.* The clear rejection of the idea of replacing the role of the teacher underlines the irreplaceable value of human action in evaluation processes.

In general, the results point to a balance between positive expectations and legitimate fears, indicating that the success of the use of *ChatGPT* in the context of evaluation depends on responsible integration, which preserves humanization, teacher autonomy, and the centrality of the teacher's role. Thus, the incorporation of technologies such as *ChatGPT* must be accompanied by formative strategies to train teachers in the conscious and ethical use of these tools, ensuring that the potential benefits are explored without compromising the fundamental values of assessment.

ACKNOWLEDGMENTS

REFERENCES

1. Bardin, L. (2015). Análise de conteúdo. Edições 70.

2. Beaulac, C., & Rosenthal, J. S. (2019). Predicting university students' academic success and major using random forests. Research in Higher Education, 60(7), 1048–1064. https://doi.org/10.1007/s11162-019-09546-y

3. Campos, L. F. A., & Lastória, L. A. C. N. (2020). Semiformação e inteligência artificial no ensino. Pro-Posições, 31, 1–12. https://doi.org/10.1590/1980-6248-2019-0137

4. Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two decades of artificial intelligence in education: Contributors, collaborations, research topics, challenges, and future directions. Educational Technology & Society, 25(1), 28–47.

5. Colônia, R., Souza, V. R., & Araújo, M. A. (2020). Inteligência artificial e avaliação da aprendizagem: Uma revisão sistemática. Revista Brasileira de Educação, 25(1), 45–62. https://doi.org/10.1590/1413-2478.2020.25.e200028

6. Cresswell, A. (2007). Conhecendo os conectores? Avaliando a aprendizagem baseada em dados em um curso de habilidades de redação. In Corpora na sala de aula de língua estrangeira (pp. 267–287). Brilho.

7. Duque, R. de C. S., et al. (2023). Formação de professores e a inteligência artificial: Desafios e perspectivas. Contribuciones a las Ciencias Sociales, 16(7), 6864–6878. https://doi.org/10.55905/revconv.16n.7-279

8. Durso, S. de O. (2024). Reflexões sobre a aplicação da inteligência artificial na educação e seus impactos para a atuação docente. Educação em Revista, 40, e2024-001. https://doi.org/10.1590/0102-469837234

9. Fernandes, A. B., et al. (2024). Inteligência artificial na avaliação de desempenho acadêmico: Desafios e oportunidades no ensino médio. Revista Ibero-Americana de Humanidades, Ciências e Educação, 10(3), 180–196. https://doi.org/10.51891/rease.v10i3.13107

10. Gardner, J., O'Leary, M., & Yuan, L. (2021). Artificial intelligence in educational assessment: 'Breakthrough? Or buncombe and ballyhoo?' Journal of Computer Assisted Learning, 37(5), 1207–1216. https://doi.org/10.1111/jcal.12577

11. González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial intelligence for student assessment: A systematic review. Applied Sciences, 11(12), 5467. https://doi.org/10.3390/app11125467

12. Hooda, M., et al. (2022). Artificial intelligence for assessment and feedback to enhance student success in higher education. Mathematical Problems in Engineering, 2022, 1–19. https://doi.org/10.1155/2022/5215722

13. Lima, T., et al. (2020). Avaliando um jogo educacional para o ensino de inteligência artificial: Qual metodologia para avaliação escolher? In Workshop sobre Educação em Computação (pp. 66–70). Sociedade Brasileira de Computação.

14. Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. Education Sciences, 13(4), 410. https://doi.org/10.3390/educsci13040410

15. Marcom, J. L. R., & Porto, A. P. T. (2023). O uso da inteligência artificial na educação com ênfase à formação docente. Revista de Ciências Humanas, 24(3). https://doi.org/10.32813/rchv24n3-2023

16. Martínez-Comesaña, M. (2023). AI in educational assessment: Opportunities and ethical considerations. Journal of Educational Technology, 58(2), 198–214.

17. Montagner, R. (2019). A revolução da IA na educação: Potencialidades e desafios. Educação e Sociedade, 40(3), 705–728. https://doi.org/10.1590/ES0101-73302019212636

18. Observatório de Educação do Grande ABC Paulista. (2024). O uso do ChatGPT relacionado às possibilidades para a educação infantil e ensino fundamental.

19. Otter, D. W., et al. (2020). A survey of the usages of deep learning for natural language processing. IEEE Transactions on Neural Networks and Learning Systems, 32(2), 604–624. https://doi.org/10.1109/TNNLS.2020.2979670

20. Preuss, E., et al. (2019). E-DUB-A: A tangible educational resource editor in inclusive classes. In IEEE 19th International Conference on Advanced Learning Technologies (ICALT) (pp. 303–307). IEEE. https://doi.org/10.1109/ICALT.2019.00094

21. Russell, S., & Norvig, P. (2016). Inteligência artificial (3rd ed.). Artmed.

22. Swiecki, Z., et al. (2022). Assessment in the age of artificial intelligence. Computers and Education: Artificial Intelligence, 3, 100075. https://doi.org/10.1016/j.caeai.2022.100075

23. Tai, J., Ajjawi, R., Boud, D., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. Higher Education, 76(3), 467–481. https://doi.org/10.1007/s10734-017-0220-3

24. Torfi, A., et al. (2020). Natural language processing advancements by deep learning: A survey. arXiv preprint, arXiv:2003.01200. https://doi.org/10.48550/arXiv.2003.01200

25. Vieira, S. (1997). Introdução à bioestatística. Elsevier Brasil.

26. Wang, X., Anwer, N., Dai, Y., & Liu, A. (2023). ChatGPT for design, manufacturing, and education. Procedia CIRP, 119, 7–14. https://doi.org/10.1016/j.procir.2023.03.002