Revista Brasileira de Informática na Educação – RBIE
Brazilian Journal of Computers in Education
(ISSN online: 2317-6121; print: 1414-5685)
http://br-ie.org/pub/index.php/rbie

Submission: 23/Aug/2024; 1st round notif.: 21/Jan/2025; New version: 31/Jan/2025; 2nd round notif.: 20/Feb/2025
Camera ready: 02/Apr/2025; Edition review: 02/Apr/2025; Available online: 27/Apr/2025; Published: 27/Apr/2025

Evaluation of Human-Artificial Hybrid Tutoring System for Mediation of Engagement in E-Learning

Avaliação de Sistema de Tutoria Híbrida Humano-Artificial para Mediação do Engajamento no Aprendizado

Evaluación del sistema de tutoría híbrido humano-artificial para la mediación de la participación en el aprendizaje electrónico

Aluisio José Pereira
Universidade Federal de Pernambuco
ORCID: 0000-0003-2960-3481
ajp3@cin.ufpe.br

Alex Sandro Gomes
Universidade Federal de Pernambuco
ORCID: 0000-0003-1499-8011
asg@cin.ufpe.br

Tiago Thompsen Primo
Universidade Federal de Pelotas
ORCID: 0000-0003-3870-097X
tiago.primo@inf.ufpel.edu.br

Abstract
This pioneering article aims to understand hybrid tutoring strategies, which combine human and artificial tutor interventions. In this regard, an approach of an Intelligent Tutoring System (ITS) was developed, involving the intervention of human tutors to mediate student engagement in e-learning. To evaluate the effectiveness of the approach, testing guidelines were adopted involving human tutors from the stages of system preparation, activity script creation and interview, test application, data collection, analysis, and communication of iteration results and proposals for improving the ITS approach. The results suggest effectiveness in increasing productivity and facilitating tutoring activities. Additionally, there was a lower perception of complexity, burden, and consistency when continuously iterating the system with the involvement of human tutors. Therefore, strategies such as "active search" for students with low engagement levels were facilitated by algorithms from the ITS approach, with the contribution and involvement of interpersonal interactions of human tutors to provide students with attractive and personalized tutoring experiences.
Keywords: *evaluation; intelligent tutoring system; mediation; human tutoring; e-learning.*

Resumo
Este artigo pioneiro tem como objetivo compreender estratégias de tutoria híbrida, que combina intervenção humana e artificial. Neste sentido, foi desenvolvida uma abordagem de Sistema de Tutoria Inteligente (STI) que envolve a atuação de tutores humanos para mediação do engajamento estudantil no aprendizado on-line. Para a avaliação da efetividade da abordagem foram adotadas diretrizes de teste com envolvimentos de tutores humanos a partir das etapas de preparação do sistema, criação de roteiro de atividades e entrevista, aplicação dos testes, coleta, análise e comunicação dos resultados das iterações e propostas de melhoria da abordagem de STI. Os resultados sugerem efetividade na possibilidade de aumentar a produtividade e facilitar as atividades de tutoria. Bem como, houve uma menor percepção de complexidade, peso e consistência ao iterar continuamente o sistema com a atuação de tutores humanos. Portanto, estratégias como a "busca ativa" dos estudantes com baixos níveis de engajamento foi facilitada por algoritmos da abordagem de STI com a contribuição e envolvimento das interações interpessoais de tutores humanos para oferecer aos estudantes experiências de tutoria atrativas e personalizadas.
Palavras-Chave: *avaliação; sistema de tutoria inteligente; mediação; tutoria humana; aprendizado on-line.*

Resumen
Este artículo pionero tiene como objetivo comprender las estrategias híbridas de tutoría, que combinan intervenciones de tutores humanos y artificiales. En este sentido, se desarrolló un enfoque de un Sistema de Tutoría Inteligente (STI), que involucró la intervención de tutores humanos para mediar el compromiso de los estudiantes en el aprendizaje en línea. Para evaluar la efectividad del enfoque, se adoptaron directrices de prueba que incluyeron la participación de tutores humanos en las etapas de preparación del sistema, creación de guiones de actividades, entrevistas, aplicación de pruebas, recolección de datos, análisis y comunicación de los resultados de

Cite as: Pereira, A. J., Gomes, A. S., & Primo, T. T. (2025). Evaluation of Human-Artificial Hybrid Tutoring System for Mediation of Engagement in E-Learning. Revista Brasileira de Informática na Educação (RBIE), vol, 33, 216-246. https://doi.org/10.5753/rbie.2025.4988

iteración, así como propuestas para mejorar el enfoque del STI. Los resultados sugieren una efectividad en el aumento de la productividad y la facilitación de las actividades de tutoría. Además, hubo una menor percepción de complejidad, carga y coherencia al iterar continuamente el sistema con la participación de tutores humanos. Por lo tanto, estrategias como la "búsqueda activa" de estudiantes con bajos niveles de compromiso fueron facilitadas por los algoritmos del enfoque STI, con la contribución e implicación de las interacciones interpersonales de los tutores humanos para proporcionar a los estudiantes experiencias de tutoría atractivas y personalizadas.
Palabras clave: *evaluación; sistema de tutoría inteligente; mediación; tutoría humana; aprendizaje en línea.*

1 Introduction

Artificial Intelligence (AI) is causing significant transformations in various contexts and domains, as emphasized by (Wang; Siau, 2019; Wamba-Taguimdje *et al.,* 2020; Abdallah *et al.,* 2020; Sadiq *et al.*, 2021; Mithas *et al.,* 2022; Xia *et al.,* 2022; Kar; Choudhary; Singh, 2022). Due to their innovative potential, technologies with Artificial Intelligence in Education (AIEd) are arousing interest due to their ability to transform the educational scenario (Srinivasa; Kurni; Saritha, 2022; Alhazmi *et al.*, 2023; Zafari *et al.*, 2023). Intelligent Tutoring Systems (ITS) are advanced examples of AIEd applications (St-Hilaire *et al.*, 2022). By integrating knowledge from areas such as Computer Science, Psychology, and Education (Figure 1), these systems reproduce educational actions (Gilbert; Dorneich, 2018; Anwar *et al.*, 2022), personalize tutoring experiences, and improve the promotion of indicators of student engagement.

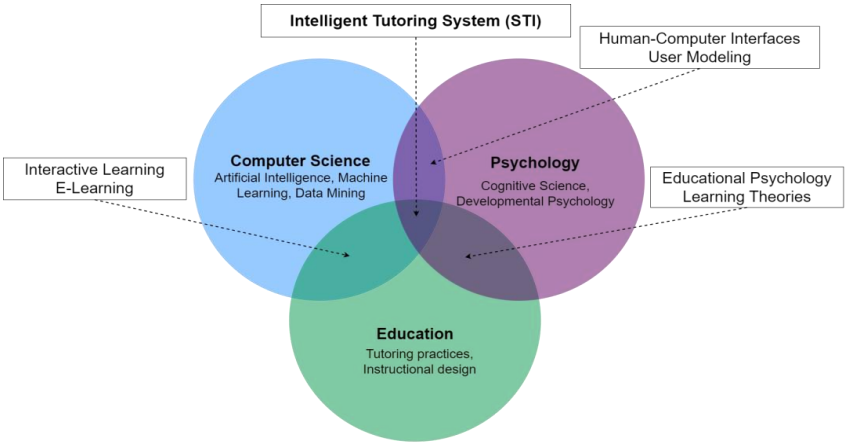


Figure 1: Coverage areas of Intelligent Tutoring Systems (ITS).

Despite advances in architectural *design* and the development of AI models adopted in ITS, replicating human tutoring styles is not a simple task. There are still uniquely human advantages, such as empathy (Maldaner; Pozzebon; Dos Santos, 2023), as well as the ability to adapt to students' individual needs (St-Hilaire *et al.*, 2022). Furthermore, there are gaps in the comprehensive system's understanding of student engagement (Phillips *et al.*, 2020), (Chen *et al.*, 2022) and (Latham, 2022). Since, to promote engagement, skills, and abilities are necessary to adequately deal with evidence of interaction indicators regarding participation, motivation, *feedback*, autonomy, performance, personalization, and among others (Ogunyemi; Quaicoe; Bauters, 2022).

That said, in the process of *designing* an ITS approach for the educational context, it is necessary to consider the still inherent role of human tutors in mediating students' engagement to be promoted. Something contemplated in the series of studies conducted (Pereira *et al.*, 2021; Pereira; Gomes; Primo, 2022; Pereira *et al.*, 2023; Pereira; Gomes; Primo, 2023a; Pereira; Gomes; Primo, 2023b; Pereira *et al.*, 2024). Therefore, this article complements this series of studies by evaluating the hybrid tutoring approach with the mediation of an ITS called "Its.Redu", in the search for answers to the following question: "*What are the perceptions of human tutors about the experiences of tutoring mediation and the effectiveness of the usability of the Its.Redu system to promote student engagement?*". The system's main interfaces are equipped with ITS strategies to support the work of human tutors and improve tutoring in teaching-learning mediated by a social learning platform context.

The article is presented in four sections: this introduction (Section 1), the background of research and related works (Section 2), the method (Section 3), results and discussions (Section 4), and considerations (Section 5). The introduction provided an overview of the topic under

study. In the method, there is a description of the techniques and procedures for collecting and analyzing tests. The results and discussions contain the findings and reflections of the study. The considerations include implications, limitations, and suggestions for future studies. It also includes information on ethical compliance and references used.

2 Background of research and related works

When analyzing ITS Strategies designed to improve tutoring for online learning students, it was observed that the United States, United Kingdom, and Canada are the most prominent countries conducting research in this field. Furthermore, ITS studies highlight three primary strategies for analyzing student engagement (Figure 2): [blue] interaction – essential for learning and development, particularly in online environments; [red] gamification – utilizing games to foster motivation and collaboration; [green] algorithms – analyzing emotions and attention to sustain engagement.

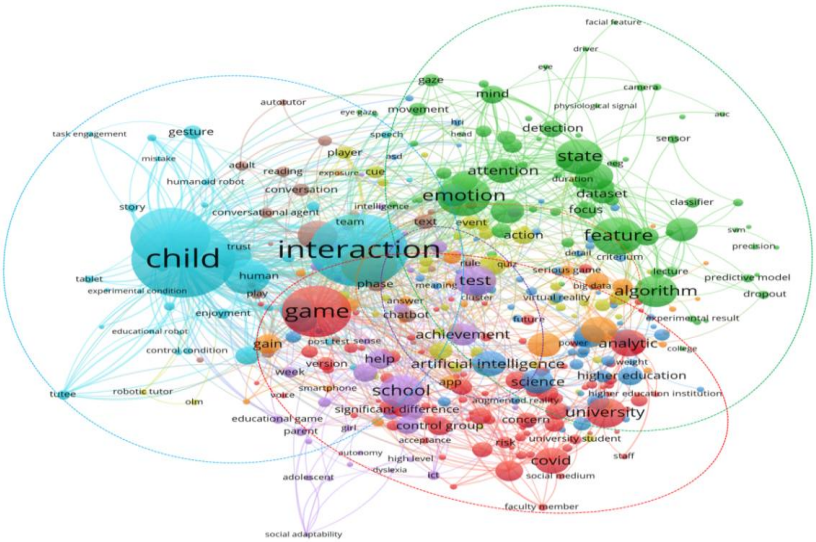


Figure 2: Intelligent Tutoring Systems (ITS) themes.

The analysis of interaction provides an indicator that uncovers numerous facets of student engagement, such as social interactions, emotional involvement, attention, gestures, confidence, and individual differences. Chen, Park, and Breazeal (2020), as well as Ashwin and Guddeti (2020), affirm that by demonstrating how diverse dimensions of interaction influence student engagement, it is possible to develop similar traits in intelligent tutors. Chrysafiadi *et al*. (2022) emphasize emotions as critical indicators that play a pivotal role in attention and, consequently, in student learning. These emotions are increasingly being explored through algorithms, such as fuzzy logic, to enhance recommendations. By analyzing student emotions — whether through expressions or dialogical interactions — aimed at mediating interactions, it becomes possible to reduce loneliness and boost motivation, especially in collaborative contexts (D'Mello; Graesser, 2012; Ruiz *et al*., 2022; Liu *et al*., 2022). According to Ruiz *et al*. (2022), student confidence can also be examined by encouraging them to share ideas and knowledge, where facial expressions and gestures can be captured to enrich the experience. Liu *et al*. (2022) further highlight that individual differences should be considered, with tutors adapting interactions to meet the specific needs of students.

However, online learning interactions often manifest in complex and varied ways, with diverse purposes and expectations. ITS approaches seek to optimize these interactions by designing different strategies, often relying on algorithms without human intervention.

Nevertheless, it is argued that using such approaches to enhance complex skills, knowledge construction, and engagement may present limitations, underscoring the importance of valuing human capabilities and maintaining human presence in the process — a perspective aligned with the studies of Baothman (2021).

Recent studies examine strategies adopted for ITS solutions aimed at enhancing student engagement in online learning (Figure 3), with a particular focus on using Support Vector Machine (SVM) algorithms (Li *et al*., 2021). Automated tutoring systems are generally categorized as conversational tutors (simulating human conversations) or autotutors (designed as a "digital twin" tailored to the learner receiving the tutoring) (Chen; Park; Breazeal, 2020).

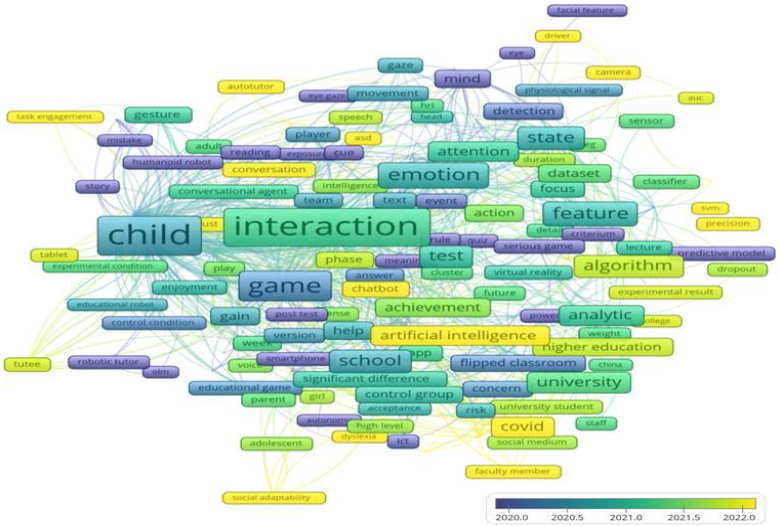


Figure 3: More recent studies (groups over the years) themes of Intelligent Tutoring Systems (ITS).

The tools employed include tablets, used to access virtual learning objects, with one study exploring narrative techniques and tutoring tips via chatbot, employing the Wizard of Oz methodology (Ruan *et al*., 2020). Cameras are also utilized to capture students' emotions through images or videos (Ruan *et al*., 2020), an approach previously explored in studies focusing on On-Task (online, virtually mediated tasks) and Off-Task activities (offline, locally mediated tasks) (Xiao; Wang, 2016), (Aslan *et al*., 2018), (Pham; Wang, 2018), (Alyuz *et al*., 2017) and (Okur *et al*., 2017). Hardware resources such as infrared, capacitive touch, and cameras facilitate communication between devices and the ITS itself, with applications extending even to evaluating students' sense of smell (Ponticorvo *et al*., 2017). These tools are particularly relevant when addressing specific student disorders (Muñoz *et al*., 2011). During the Covid-19 pandemic, ITS initiatives monitored gestures and interests to better understand students' motivation in completing activities.

The development and evaluation cycles of ITS systems aim to ensure that algorithms can emulate human-like tutoring. However, there is often a gap in collecting subjective aspects needed to inform requirements, evaluation stages, and solution design. This limits efforts to preserve and value human tutor participation throughout the project. For example, the effectiveness of machine learning models is measured using statistical indicators such as precision and accuracy. While existing studies highlight numerous strategies, there is a clear need for specific design guidelines for these systems. As Ahuja *et al*. (2022) noted, even in recent research, few studies focus on improving human-computer interactions.

This limitation can be addressed through the active participation of stakeholders from the early design stages, fostering collaboration between human tutors and ITS systems rather than seeking to replace them. Actively involving participants, combined with methodologies for

understanding the context, has been widely adopted in other research domains (Pink *et al*., 2018) and (Chen; Terken, 2022). Accordingly, the method for this study was outlined as described in Section 3.

3 Experimental Design and Study Method

For the purpose of conducting the evaluation of the hybrid (human-artificial) tutoring approach, the human role was represented by human tutors and the artificial role by the ITS system called "Its.Redu". For the evaluations, guidelines from Design Science Research (DSR) (Lacerda *et al*., 2013), and interaction design (Rogers; Sharp; Preece, 2013; Razak *et al*., 2021) were adopted. At this point, it is noteworthy that for Pimentel, Filippo, Dos Santos (2020), artifact evaluation is linked to empirical aspects of questions/hypotheses of the artifact itself and contextual conjectures that can result in technical and theoretical findings. Therefore, the evaluation of the hybrid human and "Its.Redu" (in the study, considered to be the hybrid to be tested), was characterized by the use of qualitative and quantitative approaches. In the quantitative approach, a task analysis technique was used for the tutoring activity supported by the system and developed by human tutors. Followed by a questionnaire on effectiveness indicators related to ease of use, security and trust, usefulness and weight, consistency and integration, complexity and learning. Collections occur with the involvement of the context and participants in the evaluation of the system (Section 3.1), based on the collection instruments and the procedures and techniques for data analysis (Section 3.2).

3.1 Context and participants of the "Its.Redu" system evaluations

The evaluation of the "Its.Redu" focused on the perspectives of human tutors as they interacted with students in the dynamics of tutoring in online learning. In course environments with instructional design mediated by the educational social network Redu.Digital. In this environment, tutors and students have access to digital resources that encompass characteristics of Virtual Learning Environments (VLE) and social networks (Reis, Gomes, De Souza, 2014; De Almeida *et al*., 2020), which are intended to provide classes containing Digital Educational Resources, as well as forums for collaborative exchanges and dissemination of information through the use of visual media (Pereira *et al*., 2023).

The interfaces of the "Its.Redu" system, the main artifacts to be evaluated, present different functionalities with AI strategies, related to Use Cases (UC) to support human tutors in promoting student engagement. The integration of AI into the functionalities of the "Its.Redu" system occurs through the implementation and adoption of various algorithms for anticipating tutoring demands aligned with tutoring styles (Primo *et al*., 2024a), prediction and classification (Primo; Gomes; Pereira, 2024), and the use of Natural Language Processing (NLP) algorithms (Pereira *et al*., 2024). Emerging functionalities have been identified for tutoring approaches (Primo; Gomes; Pereira, 2024a). Among these, notable strategies include monitoring student performance, identifying favorable and socially shared learning paths in online-mediated activities (Primo; Gomes; Pereira, 2025), and classifying interaction levels to enable tutors to proactively engage with students. The system specifications focus on supporting human tutors (Primo; Gomes; Pereira, 2023b) based on the contingencies of their role, challenges, and the proposal of functionalities for adopting emerging AI technologies in tutoring (Primo; Gomes; Pereira, 2023a). When accessing the "Its.Redu" [UC01] tutors are able (Figure 4) to ask questions and obtain answers about the platform and the context of the educational environment. View the tutored courses [UC03], and present the platform and learning environments [UC16], based on interactive tutoring strategies with recommendations and messages to learn about the

resources and encourage interaction in online learning consistent with the study proposed by Pereira, Gomes, and Primo (2022).

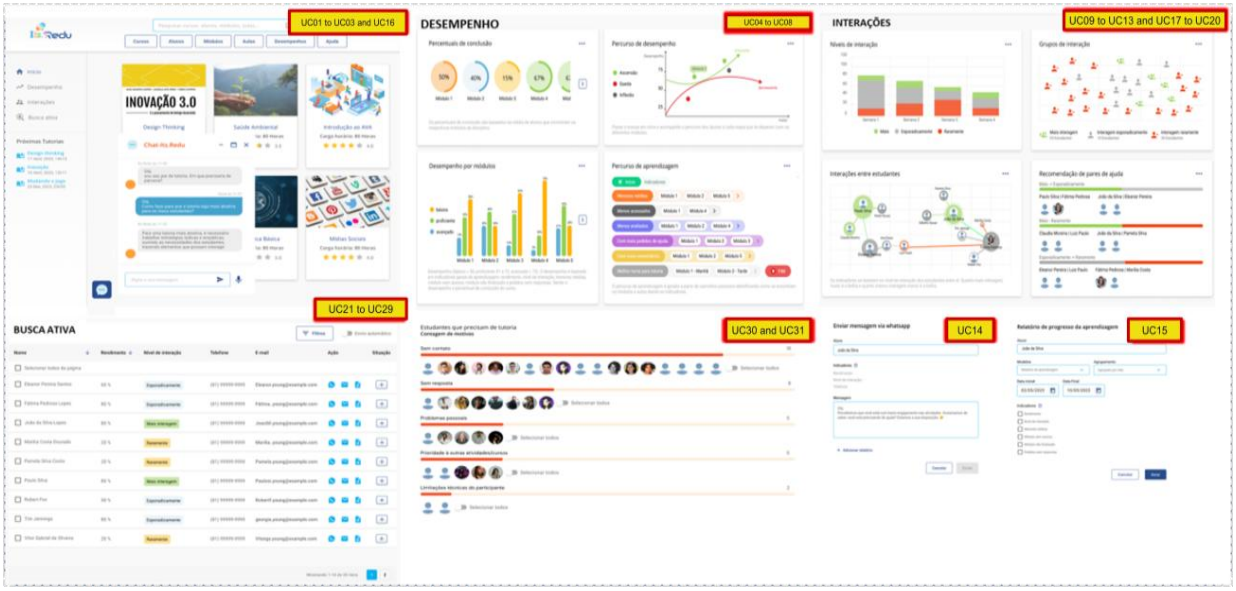


Figure 4: System interfaces “Its.Redu” to support tutoring activities.

Understanding how interactions are shared in the context of student spaces, and how learning can be shaped by individual and collective practices, provides insights into representations of student performance monitoring. This includes considering completion percentage, performance in specific modules, and performance and learning paths related to use cases [UC04 to UC08] which refer to: [UC04] and [UC05] view and monitor student performance, [UC06] consult completion percentage, [UC07] generate learning paths, [UC08] generate performance paths. The system also predicts engagement indicators through the classification of interaction levels that are associated with use cases [UC09 to UC13], which refer to: [UC09] managing groups of students by interaction level, [UC10] applying classification, [UC11] view groups of interactions, [UC12] manage learning progress, [UC13] view learning progress. And, the use cases for [UC17 to UC20], refer to [UC17] recommending help pairs, [UC18] applying collaborative filtering, [UC19] checking similarity between students, [UC20] presenting interaction paths. It facilitates the effective identification of students who need tutoring. This is associated with use cases [UC21 to UC31], which include: [UC21] view tutoring pending, [UC22] generate list of students who need tutoring, [UC23] tutor active student search, [UC24] track student progress, [UC25] check pending students, [UC26] analyze help requests, [UC27] list unanswered help requests, [UC28] track course registrations, [UC29] analyze comments, [UC30] track request tutoring, [UC31] request list of help requests. Complementary to the “active search” process is the sending of messages to students [UC14] and the generation of learning progress reports [UC15]. In addition, students who return contact from tutors are mapped to features related to use cases [UC30] follow up on tutoring requests, [UC31] request a list of help requests.

In this context, seven educational tutors participated in the evaluations (identified as T01, T02, T03, T04, T05, T06, and T07), who develop tutorials in online learning courses aimed at elementary and technical education students (associated with the commercial and industrial sector), as well as courses offered by magazines with academic publications. For Nielsen (2000), five users is a significant number to find the main problems in system testing. In demographic terms, the participants mostly had training in pedagogy or literature with an emphasis on the Portuguese language, but with notable knowledge in virtual environments and interdisciplinary skills, in addition to professional experience that varied between four and six

years in the field of tutoring in the context of online teaching-learning. The tutoring experiences and activities developed are characterized in line with digital literacy for access to educational materials and use of digital resources, monitoring of performance and interactions in the teaching-learning process, and active search for students, among others, priority outcomes for promoting student engagement.

3.2 Data collection and analysis procedures

The data collection and analysis procedures for the evaluation of the "Its.Redu" system encompassed a comprehensive set of steps. Initially, a reenactment and practical use of the system were conducted, during which tutors participated in simulated activities to assess the system's functionality, usability, and overall effectiveness (efficiency and efficacy). In this phase, both quantitative and qualitative data were gathered through the tutors' interactions with the interfaces, documenting challenges, ease of use, and insights into the user experience. Subsequently, a usability questionnaire was administered, comprising structured questions with rating scales and open-ended inquiries to measure tutors' perceptions regarding aspects such as ease of use, design, utility, and reliability of the system. The collected data were triangulated and analyzed qualitatively by identifying recurring themes and patterns in the responses, as well as quantitatively, using user interaction metrics with the interfaces and the usability questionnaire. This approach provided a comprehensive understanding of the system's effectiveness in supporting tutoring activities.

To investigate the perceptions of tutors with "Its.Redu", individual tests, were conducted with each participant. In the tests, tutors carried out activities in sections that involved examining the interactions and usability of "Its.Redu" and highlighting perceptions about support in the dynamics of promoting student engagement. The instruments used to collect the perception of tutors in evaluating the system involved an activity protocol (Table 1).

Table 1: Activities in tasks carried out by tutors in evaluating the "Its.Redu".

A#	Activities/Task
A01	Access the "Its.Redu".
A02	Freely explore the tool in up to 5 minutes.
A03	Access Module 1 of the *Design Thinking* course and report on the performance mapped by the system.
A04	Identify the interaction of the student "Fátima Pedrosa", and highlight what you understood from the classification of this interaction generated by the system.
A05	Carry out an active search for the student "João da Silva", and send him a tutoring message.
A06	Generate a system-proposed report of student learning progress.
A07	Select all students and generate a learning progress report with the system's tutoring suggestions on student performance, level of interaction, and engagement.
A08	Access the system's tutorial *chat,* and describe the experience and what you understood.
A09	Find the students who need tutoring and describe your understanding of the situation.

Note: A# - enumeration of activities/task.

In each test, tutors had the opportunity to use the system, allowing them to understand the features and report how they conducted tutoring practices without the system. When carrying out the tasks, tutors were encouraged to "think aloud", while using the "Its.Redu". In these iterations, the collection instrument was complemented with open questions that allowed the experience to be explained and answers to: "*How do you naturally carry out the tutoring activities you conducted through the system?*", "*Did you understand how the activity carried out was being supported by the system?*", and "*What implications do you understand that the support provided by the system would have on tutoring?*".

The tests conducted individually with each tutor were recorded for later analysis. In this case, the controlled experimental conditions were related to the performance of activities in each task. The records collected in multimedia files (audio and screen capture videos of system use) were used to analyze the completeness of the activities, transcription of the responses to

understand the tutors' perceptions about the difficulties in carrying out the activities, about the contingencies and possible implications in the performance of tutoring, which would allow checking the effectiveness or reorienting the interventions of the "Its.Redu" system.

In this case, the Quant UX tool (https://app.quant-ux.com) was employed to guide the testing process and collect indicators of the system's effectiveness (efficacy and efficiency). Additionally, the QDA Miner tool (https://provalisresearch.com) was utilized to process, analyze, and code the thematic content derived from the tutors' perceptions during the content analysis. At this point in the evaluations, the tangible elements experienced by human tutors were the "Its.Redu" user interfaces, based on usage indicators, in addition to the algorithms' capabilities to significantly classify and discover patterns from databases. Thus, there is the possibility of integrating these elements into usability flows that converge to support the performance of tutorials and improve student engagement indicators.

Quantitative data analyses were conducted by calculating effectiveness and usability metrics, including the average task completion time, success rate, and the usability score. Considering that, in addition to the activities, a questionnaire (Table 2) of the System Usability type was also adopted Scale (SUS) (Martins *et al.*, 2015; Gao; Kortum; Oswald, 2020; Lourenço; Carmona; Lopes, 2022), with answers on a scale from 1 (strongly disagree) to 5 (strongly agree). This questionnaire was used as an instrument to establish an interrelationship between the perceptions of tutors and indicators of effectiveness based on perceived usability.

Table 2: Usability questionnaire to test the "Its.Redu" system.

P#	Questions
P01	I think I would like to use this system often.
P02	I found this system unnecessarily complex.
P03	I found this system easy to use.
P04	I thought I would need help from a technical person to be able to use this system.
P05	I thought the various functions of this system were well integrated.
P06	I think the system has a lot of inconsistency.
P07	I imagine most people can learn to use this system quickly.
P08	I found this system very cumbersome to use.
P09	I felt very safe using the system.
P10	I needed to learn many things before I could use this system.

Note: P# - question enumeration.

The data collected through the questionnaire (Table 2, on system usage) enabled the conduction of statistical analyses to obtain responses aligned with the main question of this article by testing the following hypotheses: null hypothesis - "$\boldsymbol{H_0}$: *The support of intelligent tutoring system is* ***not*** *significant for human tutoring activities in online learning.*"; alternative hypothesis - "$\boldsymbol{H_a}$: *The support of intelligent tutoring system is significant for human tutoring activities in online learning*.". In these hypothetical cases, if we tested that the indicators were positively responded to by tutors when using the system's functionalities and identifying significant evaluations of effectiveness, it would be a sign that the system appears, in the tutors' perception, to be suitable for supporting the promotion of student engagement.

The hypotheses tested complement the perceptions of human tutors. Putting these hypotheses to the test, it is understood that the effectiveness and perception of designed artifacts can be defined by users' experiences when using human skills to manipulate technological tools and achieve desired objectives more efficiently and effectively. Data triangulation, which combines qualitative and quantitative results, enables the validation of the system's effectiveness. The results of the evaluations of the artifacts proposed for the systems explored in the digital, empirical and subjective dimensions of the effectiveness of the specifications are found in the following section (Section 4).

4 Results and Discussions

Although didactic-pedagogical theories encourage interactionist dynamics between humans and humans with the inclusion of phenomena that involve different subjects and elements of the environment and participants in the educational scenario, recently students have been having more and more contact with ITS. However, the design of these systems in short does not consider the perceptions of human tutors themselves. The proposal in this article is the design of an ITS approach with models and user interfaces aimed at supporting human tutors to conduct tutorials aimed at engaging students in online learning.

4.1 Reviews of the "Its.Redu" system interfaces

Tutors went through activity flows based on the hierarchy of interfaces of the "Its.Redu" system. Figure 5 presents the hierarchy scheme possible for tutors to follow in tests and assessments. The hierarchy is organized in depth of access, in which the system's functionalities range from the "home" screen, "chat- Its.Redu", "performance" (and sub-screens for monitoring student performance), "interactions" (and sub-screens for monitoring student interactions), even "active search" (and sub-screen for conducting active student search).

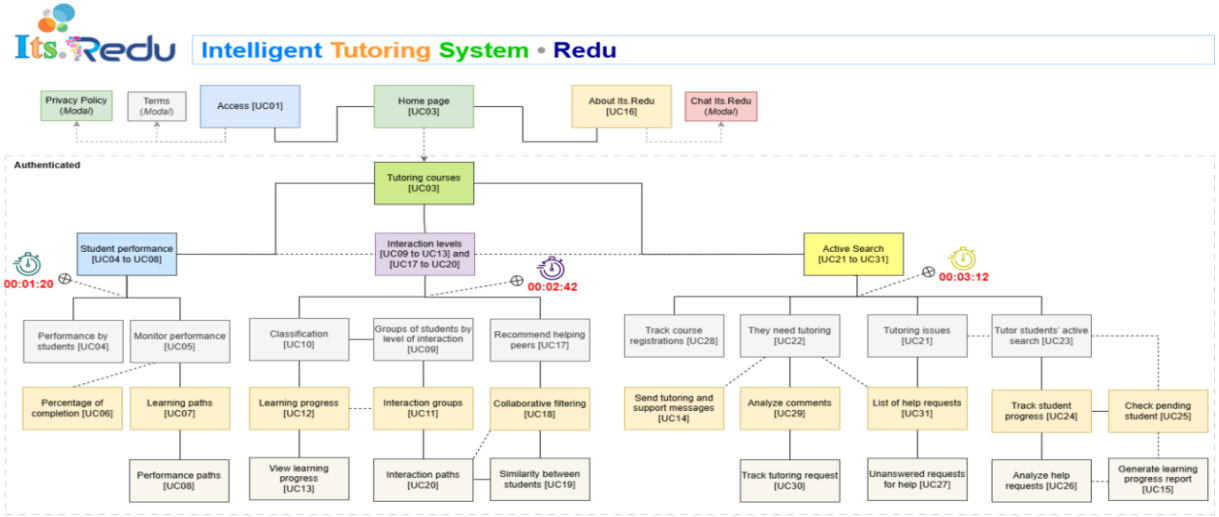


Figure 5: Interface hierarchy diagram of the system map "Its.Redu".

It also presents the average time it took tutors to carry out activities related to the depth of performance features, interactions, and active search of students. During the interface tests, it was found that all tutors were able to carry out all activities comprehensively and consistently. When analyzing the time spent in the "Its.Redu" system interfaces, an average variation was noticed, from a minimum of 140 seconds to a maximum average time of 192 seconds.

In the tutors' interactions with the interfaces, the resulting heat maps showed the usability of the screens (home, "chat-Its.Redu", performance, interactions and active search) and the mouse paths, clicks and tutors' journey (as presented in Figure 6), in which the warmer colors indicate that the majority of tutors scrolled through this part of the screen with greater intensity and permanence. The greatest permanence and intensity observed were precisely in the clickable and interactive points. The system's interfaces, encompassing functionalities related to general information, performance, interactions, and the proactive identification of students, as outlined in the listed use cases. The "highlighted" areas represent the analysis of heatmaps and click paths within the "Its.Redu" platform, revealing tutors' interaction patterns and emphasizing their interest in specific interface sections, particularly those focused on student performance and

engagement. Tutors extensively explored visualizations and student profiles, prioritizing the identification of needs and providing targeted support in their tutoring activities.

Figure 6: Mouse heat map, click, and journey of tutors when testing the "Its.Redu" interfaces.

When analyzing interactions with the interfaces, it was also identified that they varied on average from a total of screen views (initial = 28, "chat-Its.Redu" = 15, performance = 19, interactions = 25 and active search = 29), clicks on screens (home screen = 28, "chat-Its.Redu" = 14, performance = 33, interactions = 63 and active search = 91), clicks on widgets (home screen = 33, "chat-Its.Redu" = 16, performance = 29, interactions = 25 and active search = 84). A Principal Component Analysis (PCA) (Ringnér, 2008; Kurita, 2019), determined that the students' "levels of interactions" and "active search" functionalities corresponded to 73.8% of interactions with the system interfaces (Figure 7). However, it was noticed that despite the functionalities related to "levels of interactions" (with 2 minutes and 42 seconds) and "active search" (with 3 minutes and 12 seconds), they are the ones with the longest duration and with the most variations of clicks, and views from tutors. These were arranged at greater depths in the access hierarchy. In these cases, the possibilities for adjustments would be to present them in a more visible, accessible, and easy-to-locate way.

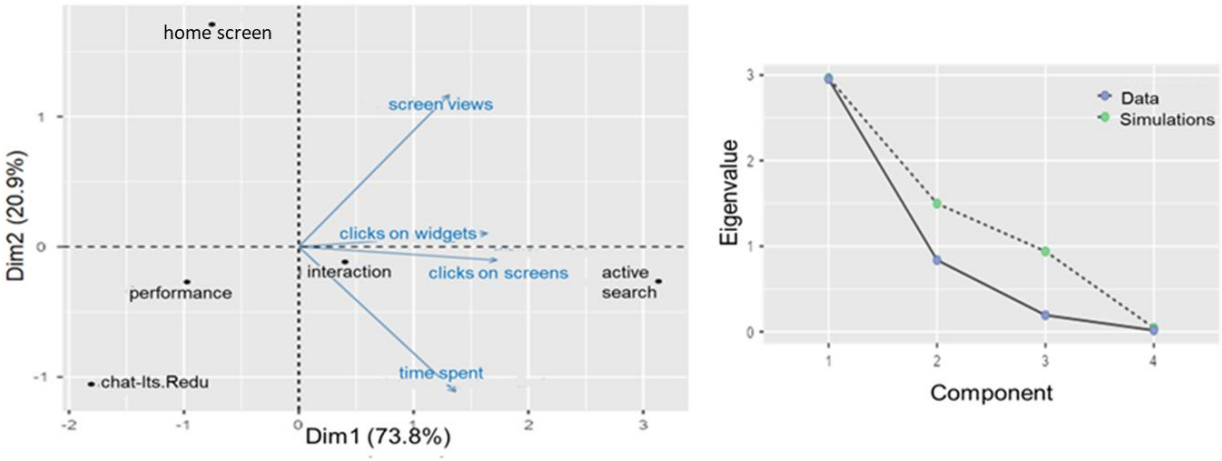


Figure 7: Main component analysis of tutors' activities interacting with "Its.Redu".

Added to these findings, even though it was noted that the tutors were able to interact with the system, the act of thinking out loud, in addition to questions asked during the activities, which complement the tutors' perceptions of the system.

4.2 Tutors' perception of the "Its.Redu"

To consider the role of tutors in evaluating the system, a position was followed that is in line with that highlighted by Conati *et al*. (2021), attentive to reorienting studies and approaches to the development of AI technologies toward emerging concerns that increase transparency, explanation, and trust in artifacts designed to carry out interventions with AI strategies in educational contexts. At this point, it is understood that tutoring practices in the teaching-learning process must value the knowledge that students already have. Furthermore, the perspective advocated by cultural psychology of education, grounded in socio-interactionism, allows us to move beyond merely encouraging students to take a stand, express themselves, and actively participate, towards fostering collaborative problem-solving. Tutors must support student engagement, even in complex and challenging situations. In this sense, participants in the evaluation engaged in activities that were designed around various action scenarios. In activity **A01** - when accessing the "Its.Redu" came across the environment aimed at the main tutoring activities conducted by them in their respective courses that they were tutors. In this case, reports suggest that the form of access is already similar to other systems used. So there were no significant considerations at this point. At the time of activity **A02** - freely exploring the system, the reports suggest situations that are mostly **receptive and promising** (underlined with our emphasis) for using the system:

> **T01**: *So, let me see. Here's the Design Think course, and there are other things. There's the menu here. Entering here has the modules for the period. Then, in this filter here, there are days, you would have the option to filter by week. I'll click here to see. I'm going to click on this other one. You have to filter by week, 15 days, in the last 7 days. Let's go back. There's a menu on the side there too. O!? Cool. This one, right? There are performances, there are interactions, here. Oh, I understand. And active search. Cool. Then I want to see people who are on low income. Or you have a low level of interaction. Then I would filter there. O!? And I would only look at those I am interested in looking at. Cool. Show. And here's one... This little thing here... Ah! Here is a chat. Very good. This chat is on every page, right? It would be standard. Yeah, that's what I realized. It's a lot.*
>
> **T02:** *[...] When I enter here, the course appears, the modules appear, a list with the modules appears, indicating the courses. The isolated modules would be the modules within each course, right? We generally consult this information. There you would have this visual indication, and the environment or each course.*
>
> **T03:** *[...] Ah! That would be the courses, then I keep accessing them. Ah, I want to look at a course. I'll take a good look. Entering the course, then within the course, there are the performances of the modules, this would be the classes, right? I understand, so I can see what's going on. If you want to change, you would have to go back all over again to be able to enter another one, right? Oh! It just needs to change here.*
>
> **T04:** *I'm going here. When you enter the course, search by module, this list appears with the modules indicating the courses, the isolated modules would be the modules within each course, right? It's good that we generally consult this information. There you have this visual indication of how the environment or each course is. Let's move on to the next ones. This would be interaction. About student activities, right? We don't have this easy. This other active search, checking. Let's go. Oh, that's good.*
>
> **T05:** *[...] I will explore here. In this first part, I have the tutor's profile. And then I have the courses that I provide as pedagogical tutoring. Oh! I have to stay on top, I have to be responsible for these courses here. I understand, these would be the courses that I have tutored, right? And I access the courses and see them here. It has performance, right? Interactions, that's cool. And, active search. There are many things. So when I, for example, click on a course there, for example, I will be able to follow these things from the students, right? I understood. Right.*
>
> **T06:** *Let's go. Checking... These are some things we basically do. There are things I do outside Redu. I speak outside Redu because I'm going to use it, I use other tools to create, like, spreadsheets, these other software products, but it's based on data generated by Redu and which I download from there. It's basically, pick it up there and then go to some other location outside. Here would be whoever is "logged in ", right, and here are the options, like, modify profile, that kind of thing, right? Okay. Here are reminders and warnings. There is also a message that we receive, right? Hmm... I found this chat button here interesting, which is very evident. Normally we don't have this. Right. Hmm... Let's go... Modules and classes, courses... Here I don't really understand why... Modules and classes up here, if they would both be included in the initial course. Courses, this logic still doesn't make much sense to me. Oh! I understand now, first I have to access a course, to access its modules. Here I am already on the course. Beauty. And then I watch the classes. So, yes, I understand here... Let's go further. Performance here would be the student's own metrics, right? Cool. Very illustrative. This is very illustrative. Interesting to see it like this. Oh! Yes. Here we filter the course. Got it, great. Now it made more sense, I hadn't seen this*

here, this course selection box. Ah, this looks really cool. I liked it, it's very personalized, well targeted, right? Interaction... Students... Okay, this is a good illustration of friends, right? The contacts... This level of interaction is here.
T07: *Let's go. The idea would be to bring the courses here, make them more visible, and allow filtering, right? Seeing here, there is filtering by module, filtering by class, in this part. On the home page here, there are ratings, these stars would be ratings given by students, right? According to the people in the materials, right? In the course, it would be like an average of the module evaluations. It would be connected, in this space, with the platform. It would be like a tutoring space, to help with activities. Manage tutoring activities. There is active search here too. We do something along these lines, but very manually and in a difficult way. Here, it seems much simpler.*

Initially, the central theme of the tutors' perceptions revolves around navigation and interaction with the "Its.Redu" system, focusing on ease of use and access, organization, arrangement of elements, and availability of tools to monitor and manage the student performance and engagement. Tutors highlight their experiences with system usability, the ability to filter and track course modules, view student interactions and use features such as chat and active searches. They discuss the intuitiveness of the interfaces and how it supports tutoring actions, although some noted the need for improvements to the navigation flow and functionality. The emphasis is on how these tools help tutors interact with students more effectively, although they sometimes require external tools for additional tasks.

A03 - When accessing the students' performance mapped by the system, some tutors highlighted that they did not understand the visualizations, by affinity, others stated that it was something that they had already been working on constantly with something similar, but with the use of data sheets and information that were, sometimes "brute" [without treatment] and needed treatment from the guardians themselves. Others emphasized more than perception and focused on actions regarding design. Suggesting approaches that would **simplify visualizations** in a more **interactive way** (underlined with our emphasis):

T01: *[...] completion percentages are based on the average of students who completed the respective modules of a course, right? Okay, give me an overview. I think it might be interesting if I could filter by student, I don't know if that's the intention, but here it gives me the general overview of my class, let's say, [...] and then I could see, I don't know, all the people enrolled, how are their performances, I think it would be interesting to have a "little filter bar". Performance path, rise, fall, modules. I believe so, too. I think all of this could also be filtered by person, in addition to being a complete class. Let me see more. Performance by modules. Okay, this part of the learning path, minors and mediums, module 1, module 2, module 5, less accessed, less evaluated, with more requests for help, with more comments, better shift. This is great, this is really good. It's cool. Beauty.* **E:** ***Now tell me the following: how do you naturally carry out this activity in tutoring, do you understand how it is being done by the system, and what implications do you think it would have on your tutoring activities?*** **T01:** *nowadays, it's difficult, right? But currently we collect this information through reports on the platform. Then we select the course. Go to management and get the conclusion report, right? Performance per module. So it turns out that this information is given to me person by person, right? I managed to create filters, anyway, knowing person by person, right? Studying by student. And it's not visual, it's just numbering. So it turns out that it's different from this one, that here it already gives me visual information. And it's even better. I think the system does something similar to what we do, but automatically, right? When we can visualize shapes, colors and performances, I think it becomes even easier to interpret. So, answering another question, I believe the implication here would be for us to have these reports in an easier way to view, right? Because sometimes a list, if we take, for example... A course that has 300 students, we will have to go out and see them one by one, right? While here, we generally have everyone. But also, right? I think there could be, again, a filter where I could see person by person, right? So it would be really cool in that sense.*
T02: *So, in performance per module. I'll go straight here. Where does it get that name? And here it comes... And I see... Performance is based on interaction levels, minors, averages, modules without access, modules... Requests without responses. Performance is the percentage of completion of the course. And beauty. So here the performance of the module. Here I have module 1. There are few students with 25% good performance. And... Many students at 50%. So, that gives you the idea. That some did not achieve good performance per modules. Let me reflect. And that's it. A large amount... You did, right? It managed to reach 75%. I think there's only one counting part left, right? If you know what the count is. Because we have to know the exact amount, because he gives me the legend of 25%, 50% and 75%. Okay, the dry and orange is 75%, the green is 50% and the blue is 25%. But it doesn't really give me the number of people who are in the situations, right? So I think it would be interested in that sense.*
T03: *Perhaps filtering the courses here, but it seems like filtering. Maybe filtering by module, where the system could help me filter by class, something in that initial part. So if there was more accessed material with good reviews, you know? Evaluations given by the students by the instructors, that the staff there in the material, you know, of the modules, so that I can understand if the students are liking it or not. In performance here, on the course. It would be like an average of the modes' evaluations, right? The system is connected to the courses. This would be like a space for tutoring and support, like this screen to see who I need to help with the activities. Very good. Because, sometimes, I use other tools to plan tutorials. I sometimes use Trello to manage some activities.*

T04: So, I think about performance also in a "*gamification" logic, the student tells the performance he wants to achieve, he will see exactly what he needs to do, and he will actively do it. So, I think it would be interesting to have a "gamification" dynamic to monitor student performance. This panel here so we can visualize performance. It was cool like that. Show.*
T05: *When I access performance, I will have the modules here. Oh! I want to see precisely this, because since the beginning of last year I have been monitoring student performance, and it is very complicated. Here we have the performance mapped and the legend about, if I want to see the performance per module, wouldn't it be this? On the tutoring platform we have to download spreadsheets, then I have to filter only the information I want, then I have to check their performance, and then I have to say – "Such and such students are performing low". – Only then can I get in touch with them.*
T06: *I don't understand this visualization here very easily.* **E: Is there a reason that you don't understand? T06:** *It's more about reading. It's difficult to read for me. What? AND... For me, like this... Where is this coordinate and the lines located... These lines here... I don't really understand the relationship between the line and the coordinate. But let me try here. Performance, quality of classes... Ah... Come on, I'll tell you what I understand. If... Here would represent the entire course, and here would be at what point in the course do I have this module. And this would be module 3, right? I think, not exactly for completion... But for activities in general. Trying to interpret it... Hmm... I think I understand. This would be, along with interaction, lower... Average... Right... This would be like a student achievement coefficient. Right? This... Lethal. It would basically be the percentage of students who are performing below 50, 25 or 75 within the module. Hmm... So, thinking about module 1, for me the performance coefficient for activities in general... For most students it would be around 75... On average 50... And the smallest portion would be in 25. That's what I understand here [...].*
T07: *So currently, in order for us to know the students' completion levels, we first go to the environment, take the performances and put them on the spreadsheet. The spreadsheet will contain the information, name, registration date, number and the course in which the person was enrolled. With this information, we go into effortful environments. Then we request a simplified environmental completion report. Here comes the complete report on all courses present in the environment. And then we look for these people who were the last ones to write on the list, to see the percentage and completion of the course. Then, with this information, we put it in our spreadsheet. And then we see who has completed it and who has not yet completed it. When they haven't completed it, we copy their name and number, take it to WhatsApp and contact them individually to offer assistance and try to understand. We basically do the same process to see if there has been any change in this level of completion. Confirm the results after contact to mark what was said or if the person did not contact us, they did not respond. And then after that, we make a report graph and start tracking what could make improvements. So, I think it's a really cool idea to gather all the information in one place, because it would make this process easier and faster. I think, but some of the visualizations presented here, they don't specifically address the information we need and probably, due to the lack of familiarity with them, they are not so easily understood. Learning path, for example, I'm still trying to absolve it. And the performance path, I think there is information that is very useful, but in this model, it is not so clear. Oh! Clicking on these dots brings more information, which makes it better. It presents some more things from students who haven't finished yet and something like that. I understand, we can go to the part where we seek access to this information and initially leave it dynamic so that it can be replicated. It's something like this that would help us, I believe [name of another tutor] does this part more.*

The central theme here is tutors' varied experiences and challenges in understanding student performance visualizations provided by the system. Some tutors struggled with interpreting the data presented visually, while others found it familiar due to their previous work with similar tools, highlighting differing levels of familiarity and comfort with performance tracking visualizations.

A04 - When identifying student interactions classified by the system, interactions are analyzed based on the number of requests for help that students make, responses received, comments made, and messages sent to tutors, other colleagues, or teachers. If you evaluate the material, if you comment on the material, if you finish the material, if you access the materials. This results in the analyzed interactions. All of this would give an extra point for interactions and classification at the most, sporadically, or rarely interacting levels. The reports suggest that it is a strategy that allows the **mapping of students** (underlined with our emphasis) who rarely interact, adding to their performance they would possibly be those most in need of tutoring:

T01: *It's cool to be able to see their level of interaction in the last 7 days, 15 days, 30 days. 30 days would be the standard, right? Then there would be 4 weeks of the 30 days here. I got it. Show. Yeah, maybe here without, right? This is obviously changeable, but perhaps having the week actually becomes more important. Understandable. I think it was a little confusing for me. Interaction groups, those that interact most, those that interact sporadically, those rarely. Cool. Here, when I click, like, most interact, the list of students who interact the most will appear. I understood. When you click, you go directly to people's profiles? So if I wanted the following, ah, I'm going to hunt down the people who are the lowest here, the level of interaction, then I would click on this list here and a list would appear with everyone who interacts the least. To do the other part of the active search, right? OK. I understood. Then, when the mapping was being done here and people were accessing it, or returning, it would become more complete with greenery. I'm understanding. Show. Interactions between students. OK nice. And peer recommendations for help. This is also pretty cool. It's cool.*

T02: *[...] here, I understand that there is a student who will interact a lot. His ball will be big, and then he will be able to interact. With people who don't interact as much, precisely to have this collaboration. But then it catches me a bit when I see a big red ball with lines pointing towards the others. So, it would be a case of adjusting to be able to understand better.*
T03: *Okay, identify the interaction of the student [student's name], right? Beauty. So, I go to the interactions part, and then I come to look for the student [student's name]. Let me see here. The student has the interaction level indicators. I understood. She is interacting with other students [students 23, student 24, student 25 and student 26]. There are others here too, 29 and 30. They are the students that I understand she is interacting with here. Student 29 and student 25 and 30, based on the colors, we can see that they rarely interact, right? Which makes me understand that red or orange is rarely, right? They are the ones who don't interact much. That's my understanding. Since it has a similar color to all other negative indicators. Which indicate that the person rarely interacts. And at the same time, we have a connection with a person who interacts a lot. A person who perhaps interacted a little and a student 13 there who – "my God in heaven, doesn't even have a 'photo' of him"–. But anyway, tiny too. So, that's what I see. This student who doesn't interact almost at all, together with the other one here, interacts with the student. Other students also call to help each other, right? That's my interpretation. Well, we don't have that. In tutoring, we don't know much about this. I find it difficult to understand these interactions.*
T04: *Okay, here in interaction. Let's go... 80% yield, so we believe she's fine. In course activities she is fine, but she doesn't interact much. That's it, she interacts little with other students. In this case, it is sporadically. That's not it. Ready. What I understand about this interaction is that it connects with this "little bond" of friendship with other students here. This one, which in turn, has these others. That interacts less. So, these are her interactions here, where she interacts with others. And maybe it's "helpful peers" for her, right? For me, when it's green, it means it's good. If there were more here in red, it would mean that it is sporadic. Very sporadic, let's say. Right. And that. What I can understand from this point, a priori, is this.* **E: *What would be the implications for your tutoring work?*** **T04:** *I think this information would still be very useful for us. Probably, the demand to look one by one would make it more complicated to understand what the student is exactly doing. I don't know if you understand exactly what I'm trying to say. But I think this information is useful in this dimension.*
T05: *I understand being very good. The system brings together a lot of the information we need most and simplifies communication, as you can make contact without having to leave the platform. And the fact that we have access to these interaction learning progress reports, directly from the student in a more agile way, means that we can better understand their situation, than if we had to, for each course, and look at every participation he has ever made. So I think that in addition to making it easier, it could help us have, I think, more informed communication. We would know more about the students, about their behavior on the platform, and this could even help us think of alternatives on how to help with student engagement difficulties.*
T06: *In our reports, despite being laborious, we have a number of interactions. We just have no way of knowing what exactly these interactions were. But we also have the quantity. So it would be something similar, right? But I don't think it would reach that level of specificity that is being shown here. I don't know exactly how the system does it. But, I believe it would be good for me to understand how the students are doing. This way, you would already be able to know the level of interaction per group among all, which has lower averages, performance. But for me, those who come first would appear as the smallest of the smallest. Then I would look at those that might be a priority.*
T07: *Student interaction levels are not something that is part of my daily activity on the platform. But here it seems like it would be good to explore. I hadn't explored something like this before. We wait longer for student demands. Maybe this would change the way, at least mine, the way I act. I understand that the system fetches this information from the platform. But I don't know what they would be. And then you create a filter, right? I think that's it. I think I answered everything, right?*

A05 - When actively searching for students and providing tutoring, the tutors' reports suggest that the strategy resembles something they did themselves. However, manual and time-consuming processes those involve selecting students, monitoring their performance and finding those who are having the most difficulty making contact outside the platform. In this case, the system allows a dynamic that facilitates some stages of the process and presents information to tutors to make decisions and **facilitate contact during tutoring** (underlined with our emphasis):

T01: I'm going to send *him a "little message". – "Hello, You may have low engagement. I would like to know if you need help."– And then I click send. Show! He was.* **E: *I got this one pretty quickly. It says how you naturally do this activity in tutoring, how you understand that it can be done there by the system and what implications do you think it would have on your role in tutoring, okay?*** **T01:** *Today we have to do that whole process of downloading reports, getting the people, the numbers, putting them into spreadsheets, that whole process which is quite massive and takes a while too. There are several stages in this process, and then we get in touch with that person, saying that their engagement is low. So it ends up taking a lot of time. Not here, it's very simple here. And then talking about the implications, I think this would optimize the time we have, precisely because we already have all the information there. I can carry out the most targeted tutoring action for this contact and also measure the level of interaction. So, I understand that the system is facilitating this, selecting performance, right? Select one or several people, in short, to be able to provide tutoring, filter and speed up this active search process. So I think that's it.*
T02: *Beauty. Here you can test and "*link*" with the interaction and performance badges there, right? Ah, that's another possibility. I hadn't thought of it that way, either, but I think it's a cool advance. Cool. This panel here looks really cool. The visualization is much simpler. I've already done some active searches, it's quite difficult and massive. But here everything seems much easier. I did it with spreadsheets. And then in the spreadsheet we have to filter the period we want, all of that. It is much more manual and the visualization is really just the cells and text. There's no way we can have the level of*

interaction already classified here, it's... So, the interface for us to view in the spreadsheet is not as simple as it is here. It's very apparent here.
T03: *I think it was very easy and intuitive. I didn't have any problems at all, no. Give you an idea of which students are having difficulties and in which modules intervention is necessary. Because students are performing lower in this module. And I think the tutorials are more personalized and can be targeted at subjects that are more difficult, or because the student is having a problem.*
T04: *We do this with spreadsheets after checking performance and from then on we get in touch via WhatsApp to find out the reasons for those who were unable to complete the course activities. Why they couldn't do it, and we'll classify it on the spreadsheet to understand that too. Here it is, it's much more direct. This facilitates some mentoring intervention. And it makes the data analysis we do much faster.*
T05: *[...] So, I have to get in touch with him, then the idea is to bring some of these difficulties or very repetitive activities here and say the following, you need tutoring here. And then that's just the most important part of looking at those who need it. If you have this assessment of performance, of interactions, it becomes easier, I think.* **E:** ***How do you naturally carry out this activity of looking for a student and sending him a message*** *?* **T05:** *Either can come upon request. Specific, right? From a student who is asking for help. The more normally we do the following. With student data, performance something like this. We know for each person [student] what their level of completion was. And we filter those who are below 70%. And we will have this list. All students who are below 70%. If they are below 70%, they have not completed. So it's important for us to understand why it was. We go to WhatsApp and get in touch. And we write a standard message. Like this one that's here in the system — "We noticed that he has low engagement. In this case, he would need help"—. Or you enrolled but didn't finish. Always need help with something. Do you have any specific questions? Then we go there, clicking number by number. And start talking. Then paste the message. And wait for the answer. It's a massive activity, and it's all done very manually. Here it seems to be more dynamic and faster.* **E:** ***Understood. So, here I ask you another question. What implications do you understand it would have on your performance?*** **T05:** *The way it is being done here in the system. Ah, I think that mainly because you don't need to generate everything on the platform. And another is to go student by student, get in touch, get a message, copy, paste. This is more direct and without leaving the same environment. A suggestion here would be about learning progress, taking it individually or by many, or just in the next contacts being on our own.*
T06: *[...] other tutors get in touch, then, when we get to the next contact, the students say – "no, I already got in touch!"–, or sometimes you're lying, or sometimes you've actually accessed it, right? Then, in that part of searching, there is a part, a sensation, very delicate for this active search. Like, it would already bring this income to us without having to download spreadsheets, and it would bring a level of interaction. And let us understand, okay, he interacts more, they rarely interact, and get in touch. One thing is that contact was made, right? Sometimes, the person does not provide this, it is their right, not to provide the contact, and then, by contacting them, a report would be generated to add a reason, or if they wanted any of these reasons here. It's like this, right? And then say – "Ah, I got in touch, but the person has problems. I got in touch, but she has technical limitations"– Then tutoring comes into these cases, and then I will help, instead of having to waste time looking for each one.*
T07: *Here in the active search part, this is very interesting. We need it a lot. The fact that access is now easier because currently, the process is for us to take this data and look at the students, transfer it to a spreadsheet with the name information and what is needed, for example, the telephone number. And from this spreadsheet, we take it to WhatsApp and manually enter, right, create a conversation with each person to get in touch. It gets hard. This would make sending easier, hence the add report messages that are generated, it could be individually sending a report of it, or it could also be those reports that I mentioned, right? The other [name of tutor] who is doing this part more. And then these indicators would be updated according to the information in the system. It's good like that.*

A06 - When generating a report on students' learning progress, tutors highlighted that they would be more assertive, tutoring would be easier, and they would save time on tutoring activities. At this point, effectiveness in completing the task is understood as the ability to understand student metrics more clearly, and the possibility of streamlining the removal or **choice of student performance** (underlined with our emphasis), performance and interactions metrics is suggested:

T01: *So, normally, we need to go from course to generate this report. And then, it takes time again for that, right? These are reports that are not easy to download, in the sense that I can't keep, okay, I can't keep downloading this report every hour. So, but that doesn't mean that hourly, it can also be out of date. So, it ends up being a bit of work, right? Sometimes, I want simple information and I need to delete it. Anyway, so, in general, today it continues to be done in a very massive way and then, the way the system is doing it, it is very interesting, because I can understand more things, and also select only the students I want. So, for example, if there is someone there who already has a good interaction, with good performance, I don't need to worry too much about him at that moment, so I'll just understand those that give me a certain headache, and then I can have this generated report, right? Here, in a simpler and more objective way. And that's really it, in terms of having better efficiency at work.*
T02: *I come in filters. Select everyone. And then I will generate the report. Ah, now it was a learning report. It is grouped by month. Here's the month, right. You can enter income, interaction levels, averages. Oh, you can add them all, right? Ok. I clicked on generate. Would it be this. Then I printed it.* **E:** ***Was this really quick? As you naturally do, this activity in your tutorials. Do you understand how the system is being created and what implications you believe it would have for tutoring?*** **T02:** *So we use reports with graphs that we generate ourselves. But nothing like this, with similar tips here. Related to this here. I see this could help. It might be one thing if the report could have a template to choose from. Because there are some reports that we generate to send to other people who want them, coordinators, teachers, to people.*

Sometimes we generate it manually. And so yes. It would be something more targeted and specific. Like, Oh! I want this, I want that.
T03: *Ah, report. Cool. Just look. When you select here world, oh !? It is very similar to the type of information that we consult a lot, because we need to monitor the student's performance and this is information that only the report can offer us a good idea of how the student is progressing through the courses. But, it is a process that is quite time-consuming, because we have to write one report at a time and that is if it is not updated information. The system does not allow us to consult data more directly. And I think that the fact that we can't see all the reports at once ends up hurting us because we can't have a complete view of the student's situation at the moment. We are trying to identify who are the students who need help, who are the students who have already completed the course or who are the students who are underperforming, we were unable to get a quick view of who they are, and then we end up having to consult each report individually and this ends up making things easier, especially when it comes to spending a lot of time so as not to interfere with our work.*
T04: *I understand that this part of the report has implications in terms of not having to generate it on the platform. I don't really know why I didn't get to do this part much. I haven't gotten around to looking for this tool yet. But I understand that it would save a lot of time and I think assertiveness too because as we go through different situations we end up losing some information. And because a lot of time passes, sometimes the parts that I have usually become out of date. So, maybe we have this more obvious data [evident] so it would be easier to check and not end up wasting time. So the level of assertiveness and time savings are certainly what I see as implications.*
T05: *I think so. A priori, a first point would be the layout here. We click and go straight to the data. Today. Entering the course and having to monitor each student individually is a lot. And so, I think, it saves an interesting amount of time. We can just click here and it opens directly. So that's the first implication I think. The second, that already having all this data in the form of visualizations that we need and cannot get directly on the platform, makes things easier to monitor student engagement. And understanding the metrics, I think it's easier that way. Because it's visual.*
T06: *Here we sometimes use this approach of generating reports to see student performance and then carrying out active searches. It's just more complicated. To accompany students in a personalized way. And tell him if he needs help with anything. Or show the report to others in meetings. Coordinators, teachers, and staff. To take measures, you know? What was his motivation for not being able to complete it yet? These things.*
T07: *That's precisely the complication. Having to keep generating reports and taking it part by part. Not a standard that we can go straight to and understand the student's needs. You have to keep downloading the spreadsheet to get everything I want in the performance report, what I want in the report. Here I realized that the idea would be to automatically generate these parts for us. And I can see that later I can send it to whoever I want, for coordination, for the student, for the teacher, right? One thing I see would be perhaps understanding courses by city as well. We need to know where the student is talking about, what city, what course something like this was missing. But overall, it's beautiful.*

A07 - By having the learning progress report with tutoring suggestions, it was possible to understand, based on their reports, the analyses that the tutors carried out to streamline the collection of indicators such as comments, requests for help, interactions, and student performance. And with this, **suggest mentoring directions** (underlined with our emphasis):

T01: *Okay. Already understanding how I see how students interact today. There is no system or functionality that allows me to actually see this. Filtering, generating reports, or something along those lines. So, today, we can't have this type of analysis, unless I go content by content, look person by person, comment by comment, to see what the interactions are. But I don't think that has anything to do with this report, no, sorry, with this type of visualization. I think they are totally different things. So, today, we don't have it. It would be very interesting for us to be able to see how people's interactions are going, to see if they engage, if they can help each other. So, I think it's interesting in that sense. And the possible implications are precisely being able to generate collaboration networks. So, if I know which guy he interacts with a lot, he can connect with someone else. I can, as a tutor, as a support person, generate this knowledge so that these people can collaborate. So, I think it would be interested in that sense.*
T02: *It was good like that. You can even merge performance with interactions. It was always something I thought of, a type of visualization that would be the student's name, course and that focused on the interaction along with their performance. It was good to have a report here that I could achieve this. Here we can already make this option, right, ready for the interactions contemplated here. I thought it was a good one.*
T03: *I understand that I can do it if I just want the income. I just want the interactions. Or you want both. I only miss one thing, which was a report per course and from this course I managed to get all the students, you know, this one I can get the student's report, that's what I understand, is that I can get the performance report of the student, I would also think about the possibility of a report per course.*
T04: *From the report I saw everyone there, I understood. I understood. If I were to change something. I would put it here, in these selections, right? And if I wanted to change it, I would put it here, you can specify it, right? For example, I only want students who are less frequent, for example, the level of performance, then the level of interaction, I only want those who are interacting less that I want to carry out interventions to send messages to.*
T05: *We get four types of reports if I'm not wrong. On the platform, we have one about courses, subjects and modules. But there is no way for us to take these individual reports and understand the information like that. We have to filter the information within these reports, so it is very difficult for us to have and find what we need. It seems to be quite simple here. And there it means that we don't have a complete idea of all student interactions per period like we might have here. To be able to understand how his progress has been, the conclusion is that I think this way we could better monitor the students' performance. I believe they can help to find out which courses they are having the most difficulty with. What are the modules that they are not able to advance to have something specific to the module if — "Pull this learning difficulty" — I think it*

would be a more personalized view of each student if you could generate a report again just for all students in a Once a report of many interactions, of those with the lowest averages of all students, is very quick in other contexts, there is the option of generating individually or two three four different ones, or for everyone, right? That's quite interesting.
T06: *From the report, I can get things [information] that I normally can't follow. Even more so if I go through the tutorials one by one. There are moments for individual conversations, with a specific student. But you need to demand first to understand what is happening to reach the student who needs tutoring. And that's when I can understand, talking to him, why he's having difficulty and why. Here it already appears in a list, right? A list of students, I can see the list, but then this file here seems to be after this next marking there is this marking option which is one of the categories that the girls use, for example, they got in touch and then the person has a problem. Ah, she has low engagement because she has problems.*
T07: *By having the report here. The idea would be, like, automatically generalize that part of the reports that we talked about. One of the things that, even these days, we mentioned was that we needed to download the spreadsheets, we had no way of accessing this online, so that part is not enough, but the interface also allows us to select what we want it to be cool. Just one point regarding the model, there are more types and models of reports that we generate. There's performance, learning reports, but here we have some insights, that's good. To know about people [students]. The main thing would actually be a performance report on utilization by modules, so there is already a report here. So, it's good to visualize it like this.*

A08 - When accessing interactions in the tutoring chat with the support of the system, it was noticed the possibility of **automating some difficulties** (underlined with our emphasis) that are most common in tutoring, or of responding to requests for help, mainly because they do not have anything similar to what is proposed integrated into platforms educational:

T01: *So, I go here, in the chat, on this little blue ball, and there are commands to help me with the tutorials. It would be if I need a partnership in tutoring. How can I make tutoring more attractive? Then it is necessary to work on empathetic strategies, I think it is very interesting, right? If we are talking about a chat, that will provide support to the tutor. I think this is great, right, because if I have any doubts, oh, how do I do this, how do I do that, maybe for me directly or for the other tutors who work in tutoring, I have to think about the tutor who is working in the courses. It's very interesting, right, because then we can even optimize tutoring time, with something like this. I already have this automatic support to help, so, overall, I think it's very interesting. And, answering the questions, today we don't have things like that, with chat, of this type, directed at the tutor, I've never even seen if people care so much about environments aimed at the performance of tutors, our performance — "God protects us, right?"— It's a way of helping us, created for ourselves, it's really cool. We still do a lot of things manually, so we're the ones who answer questions and everything, right? I think it would be very interesting to have this help to respond and help students. I understand that it [the chat- Its.Redu] is automated, answering some questions that can be resolved through intelligence, right? I think the possible implications would be to solve problems more quickly, so that we can provide support to students, then human support will be available for situations in which they really need to act.*
T02: *Let's see. This would be a chat, right? To talk to an intelligent tutor, in this case. That. Cool. It was therefore very easy to access and interact. Look, there would be answers, right? Very good.*
T03: *You can also ask questions here, right? So the system would be something that would give us a clearer idea of things, performance, student interactions and a quicker idea of their situation to be able to offer more effective assistance [tutoring].*
T04: *Oh, there's the chat here too. Would it be an assistant for us or for the students?* ***E: for tutors.*** **T04:** *So, when I interact here, my doubts are answered. How to improve at tutoring. It would be simulating doubts that we have. It seems like something interesting. It, would be standardized for us. Then it would be more targeted. It ends up being something that can help us too.*
T05: *So, this would be a more automated way of interacting with the system, right? I've been thinking about this. I always think about the possibility of having something more focused on assistance [tutoring] for students. I, think it's possible to have interactions, but I think it's always necessary, on the human side, right? Collecting information, something we talk to students personally, making drafts of difficulties. Generally, conversations always involve a lot of information with students, some parts of which it is not possible to know everything, I think. The idea is to think of something that gets closer to the student, right? I don't know, maybe chat like this would be more distant. I'll be thinking here to see if I have any ideas to improve.*
T06: *I think this way we can understand what is happening to more students, send messages to more people. This [the chat] is to help us too, right? It can help us with different questions, from what I understand. It can help with active search if we have difficulty doing this too. I thought it was cool in terms of improving the way we currently do things. I see that this can greatly improve the platform. Mainly because we don't have anything similar.*
T07: *This chat will make it possible to deal with the things we sometimes use. Sometimes we use others like ChatGPT, which we already use for some things. But, this would just be ours on the platform. It's well aligned with what we do. Some are more generic. Here it seems to be directed towards our things, right? So, normally according to the tutoring demands. We don't have anything similar, it's more out of necessity. Here would be a good one.*

The description highlights that tutors realized the potential for automating common difficulties and responding to requests for help. The mention of automation of responses and support in interactions suggests that the system contributes to efficiency and agility in meeting students' needs, which can improve engagement.

A09 - When finding students who needed tutoring, the reports detail that, in situations without the system, if there was any change in performance during the time they used data to carry out tutoring, they would already have outdated information. However, with the system, they would have a temporal understanding, at the moment they needed to be classified according to types of difficulties already recurring in **student engagement** (underlined with our emphasis). This allows you to direct:

T01: *Ah, so I believe that the two things complement each other, right? This way, the reasons why they are not accessing will appear here, so I can create an understanding of the difficulties in a more targeted way. This is very good, so I can adjust my way of tutoring. It's great to see it like this. Because we have to keep asking questions and there are no answers, and we don't know. That's how it goes with their answers, right? I myself monitor some things to see if students are having difficulties. This way is better.*
T0 2: *This is where students have difficulties. Very good. Some so out of touch. Then filter those who are already experiencing technical difficulties, personal problems, without contact. Then you can analyze, or automatically mark people who have not responded here. Then, here, it would be to send a message on the tutoring page. It can be done via WhatsApp, email... Here, right? What is this one from? That third one? Which? That third icon.* **E:** ***to change the student's status, after conducting a tutorial.*** **T02:** *Understood. Then he will change his situation from this corner here to others, right? Cool.*
T03: *We wait more for calls. Through calls, we receive those who have doubts and those who need help. Together with the other girls [other tutors] we keep tables of who asked for help and who responded and when they responded. If it was a technical question, or other questions. Then we help more like this. When we find someone in need, we help. But here, according to the system, this initial part would already be completely done, right? It would be more up to us to provide assistance [tutoring] to clear up doubts. Seeing here, it is already classified similar to what we do as we perceive it. That's how good it is to act. I will love using the system and doing things like this.*
T04: *We generally have some personal problems, technical limitations, in using the platform. I even think that some are because students don't usually use learning environment platforms. The first thing I do is find those who need assistance [tutoring]. Generally either by report if it is something personalized or by calling us. But I think this has already been addressed and addressed the limitations that students have. Here I could have just this contact with them and then get in touch again if I needed to try to resolve the doubts. In tutorials or, if applicable, via phone call or email. Beauty.*
T05: *We keep looking for people who are not finishing. Then, we take a sufficient percentage of completion or not, we copy the name and telephone number of each of them, create a new spreadsheet, take it to WhatsApp and get in touch with them to try to understand why they need it of tutoring. Just to help. We collect it one by one, bring it, filter the information and then look and get in touch, environment by environment. One of the situations is that until you can filter the information, get in touch and then, then, it goes on, ten, fifteen days, seven days, a week, two weeks. So, when in a way, when a stage ends, when another stage starts again, it's already kind of outdated in terms of what's already there. Sometimes, it was even some of the conversations we had.*
T06: *For us to reach the students, it would be through the calls we receive. And we would manually tabulate it in a table, classifying what type of difficulty we receive. So let's go, for example, one day we received 10 calls, we would put it there, a call about access difficulties. And some, we have to ask other people for help. Others we have to resolve ourselves. I don't think we would exactly be able to know this from a report. With a performance report or an interaction report... It just seems like you can somehow tell, right? From student comments. With direct contact or more comments, sometimes you can tell.*
T07: *For me, all the system options would be great. It would be great if we had the option to generate a collective report for all students or to select a group of students and generate a report just for them, then we would have a clearer and faster view of what is happening with the students. I think we would be able to offer more qualified, more personalized assistance [tutoring] for students, then we would be able to better understand their situation and what is happening, with each one individually. This would be great for our day-to-day operations.*

The Figure 8 presents the themes based on the different perceptions of the tutors. The theme "**Receptive and Promising**" addresses the organization of modules and navigation within educational systems through intuitive filters, enabling tutors to explore courses in a visual and accessible manner. Performance and interaction are central concepts, with an emphasis on active search and interactive menus that enhance learning. "**Interactive Way**" highlights the importance of reports and visual data to monitor students' performance. The proposal includes tools to filter information, understand module completion, and provide an overview of class progress, facilitating decision-making in educational settings. "**Mapping of Students**" focuses on analyzing interactions among students, mapping groups and levels of participation. It emphasizes identifying profiles and recommendations to better understand engagement, offering easier access to visual data and mapping sporadic patterns for tailored interventions. "**Facilitate Contact During Tutoring**" explores ways to optimize communication with students by integrating direct contact via messages and reports. Using filters, the system simplifies the

identification of low performance, engagement, and interaction, enabling more targeted interventions in the tutoring process.

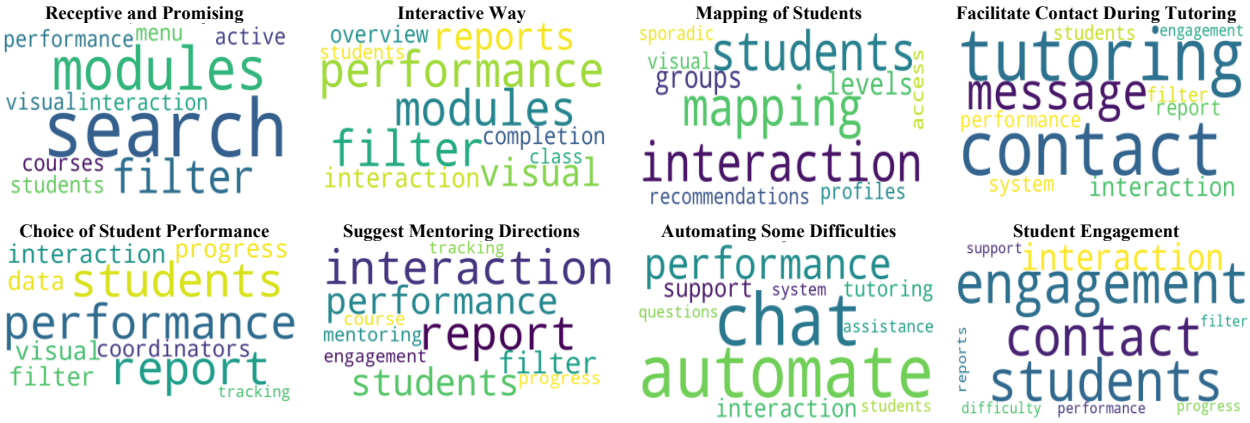


Figure 8: Word cloud by theme of tutors' perceptions when carrying out each activity in the "Its.Redu" system tests.

The theme "**Choice of Student Performance**" examines detailed reports to track students' performance. Through filters and clear visualizations, coordinators can monitor progress and efficiently perform tracking, ensuring the identification of specific needs. "**Suggest Mentoring Directions**" proposes combining interaction and performance to strategically guide mentoring efforts. With personalized reports and intuitive filters, it becomes possible to identify students with low engagement and adapt efforts to the individual progress of each course. "**Automating Some Difficulties**" emphasizes the use of automated chat and intelligent support for tutoring. Automation of tasks such as data collection and answering questions helps optimize support for students, improving interaction and allowing tutors to focus on more complex challenges. Finally, "**Student Engagement**" explores tools to identify and address engagement difficulties. By focusing on reports and direct contact, tutors can analyze factors impacting performance and interaction, providing personalized support to improve students' progress.

Despite the "Its.Redu" provides intelligent agent strategies to encourage some automated processes so that the most engaged students can cooperate with other students, without overriding their behavioral structures, disseminating examples of success by allowing behaviors in the same way as the most engaged students. It is important to understand the fundamental role of human tutors, especially when dealing with indicators that manifest themselves in various aspects and are present in physical, digital, and subjective environments in students' cognitions in online learning. The perceptions collected allowed us to summarize in Table 3 the possible benefits of the system, as well as the limitations of its use to promote student engagement.

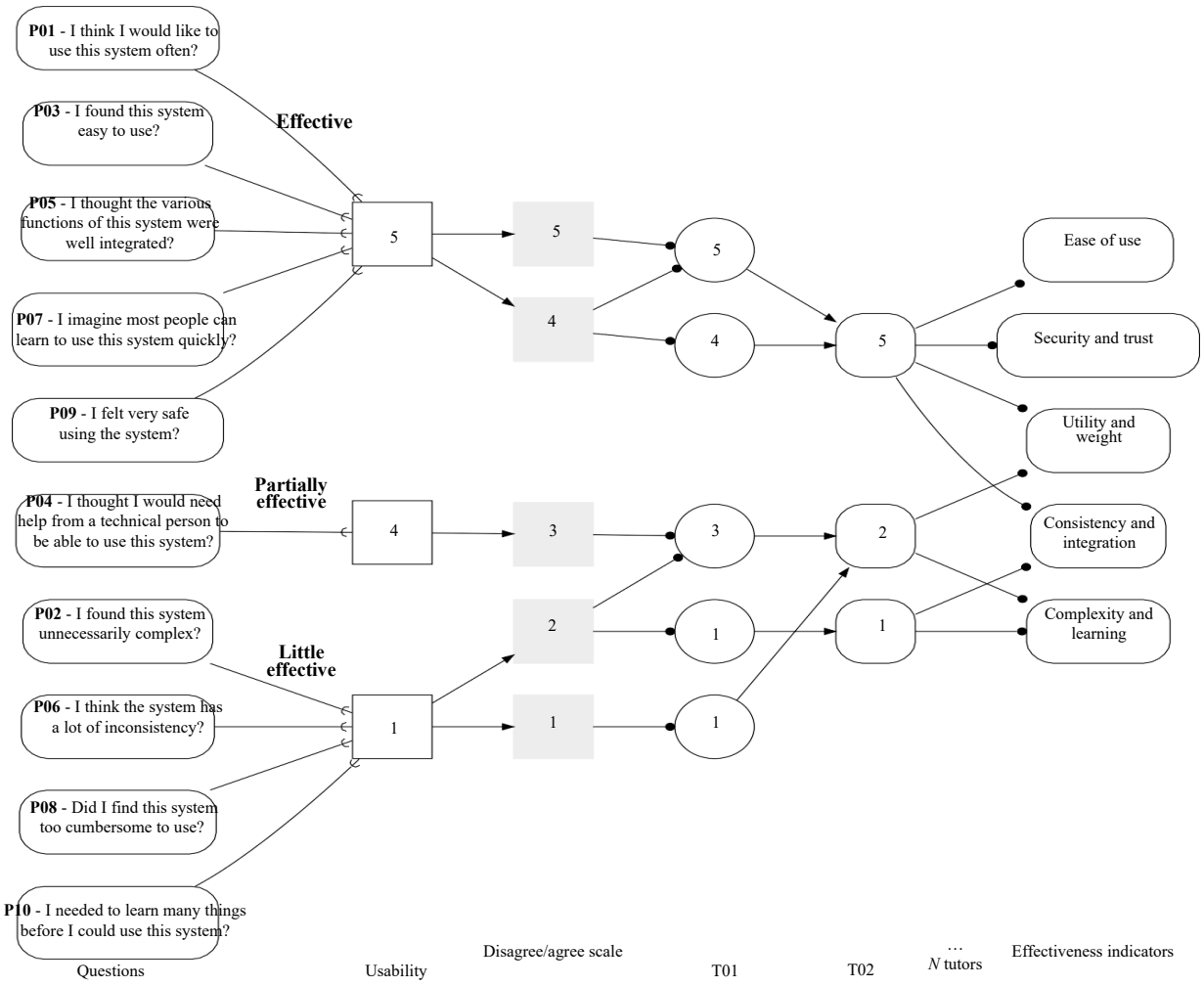
Table 3: Benefits and limitations of using the "Its.Redu" in promoting student engagement.

Benefits highlighted	**Limitations highlighted**
The "Its.Redu" provides predictions about student engagement levels. To allow personalized guidance for tutors to work with tutoring strategies and styles aimed at improving engagement according to students' types of needs.	Tutors may initially feel interested in the system, but may lose interest due to changes in the instructional *design* of courses or contexts of use.
The "Its.Redu" supports tutors in actively searching for students who have low engagement indicators.	The "Its.Redu" provides discovery of patterns in student data, but it constantly needs details and information about students, which must be fed by another platform and by tutors.
The "Its.Redu" allows targeted interactions, depending on tutoring styles, to promote student engagement.	The "Its.Redu" is interactive, but subject to predefined responses. If new types of responses are necessary, it is necessary to (re)train the models.
The "Its.Redu" features dynamic visualizations of engagement indicator patterns rankings and discoveries.	The "Its.Redu" is linked to the tutoring context. This context may vary depending on new student interactions. This may reduce effectiveness.

Reports suggest the possibility that the “Its.Redu” serves as support for human tutors, incorporating a variety of functionalities enriched with AI resources. Something similar to what was proposed Echeverria *et al.* (2023), with hybrid AI strategies with human assistance. However, it is known that the effectiveness of the system can be much more complex than the perception of the system's possibilities. At this point, different indicators were analyzed to compare perceptions of usability and effectiveness of the system.

4.3 System effectiveness indicators “Its.Redu”

The analyzes that followed involved indicators, extracted from the questions asked in the questionnaire, with the agreement/disagreement scale, tutors' responses, and effectiveness indicators (as shown in the diagram Figure 9). The Figure 9 presents the relationship between different evaluation factors (such as “utility and weight”, “complexity and learning”, “ease of use”, “consistency and integration”, and “security and trust”) and their perceived effectiveness. The diagram on the left visualizes the connections between participants and the criteria being evaluated, while the graph on the right categorizes effectiveness into three levels: “Effective”, “Partially effective”, and “Little effective”. The colored connections indicate how each criterion influences the perception of effectiveness, highlighting which factors are most associated with positive or negative evaluations. In the following situations, we tested whether the system “Its.Redu” presented: **effective -** high average score on the indicators (ease of use, security, trust, utility, and weight), indicating that the majority of participants considered the system easy to use, integrated, safe and useful. And average in the indicators (consistency and integration), indicating that participants did not experience difficulty in understanding the consistency of the system. As were also low indicators (complexity and learning), indicating that participants did not experience difficulty using the system; **partially effective -** average score on indicators (ease of use, security, trust, consistency, and integration), indicating that some participants found the system easy to use and useful. Low scores were obtained in the indicators (consistency and integration, complexity, and learning), indicating that participants experienced difficulty using the system; **little effective -** low average score on the indicators (ease of use, security, trust, utility, and weight), indicating that some participants had difficulties using the system, found it complex, inconsistent or not very useful. And a high average in the indicators (ease of use, security, trust, utility, and weight), indicating that participants did not find it easy, safe, or confident to use the system.



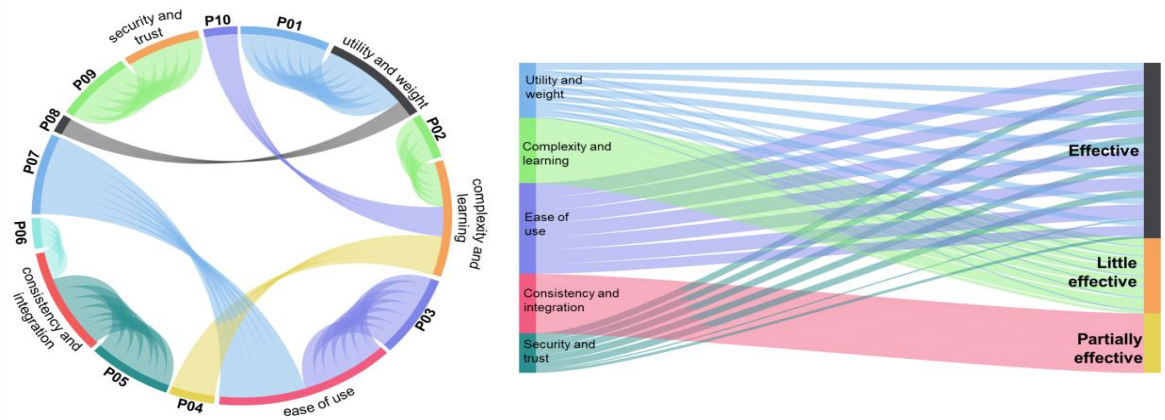


Figure 9: Diagram of the indicators analyses regarding the effectiveness of the “Its.Redu” system.

When analyzing the indicators (complexity and learning, consistency and integration, ease of use, security and trust, utility and weight) comparatively, the results indicate that ease of use received the highest ratings (Figure 10). Mainly, about how easy the tutors found the system to use, and how much they realized that most people could learn to use the system quickly.

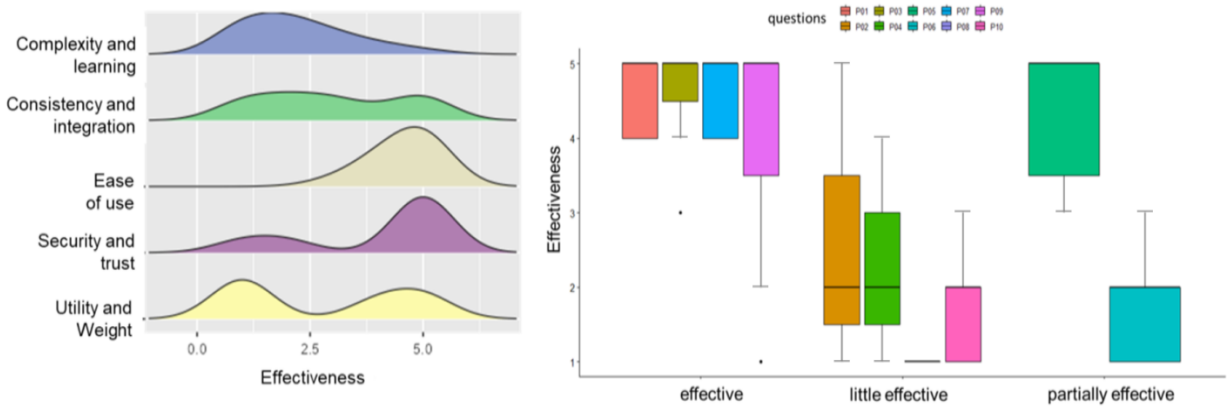


Figure 10: Density and frequency in evaluations of the effectiveness indicators of the "Its.Redu" system.

When considering the indicators on the effectiveness of the "Its.Redu", it was confirmed that "ease of use" was the indicator that obtained the highest average ($\mu = 4.57$). The "complexity and learning" indicator, regarding the need for learning to use the system, was the one with the lowest average ($\mu = 2.19$) in the evaluations (Table 4). However, these results alone are not enough to confirm whether there is a statistically significant difference between the indicators to validate the effectiveness of the system and test the **hypotheses raised**.

Table 4: Descriptive analysis of the effectiveness indicators of the "Its.Redu".

								Shapiro-Wilk		Percentages		
Effectiveness indicators	***μ***	***Me***	***σ***	***σ²***	***IQR***	***Min.***	***Max.***	***W***	***P***	**25th**	**50th**	**75th**
Complexity and learning	2.19	2.00	1.17	1.37	2.00	1	5	0.864	0.008	1.00	2.00	3.00
Consistency and integration	3.00	3.00	1.57	2.47	2.75	1	5	0.869	0.040	2.00	3.00	4.75
Ease of use	4.57	5.00	0.65	0.42	1.00	3	5	0.688	< 0.001	4.00	5.00	5.00
Security and trust	4.00	5.00	1.73	3.00	1.50	1	5	0.643	< 0.001	3.50	5.00	5.00
Utility and weight	2.79	2.50	1.89	3.57	3.75	1	5	0.724	< 0.001	1.00	2.50	4.75

Note: μ - Average; *Me* - Median; σ - Standard deviation; σ^2 - Variance; *IQR* - Interquartile range; Min. - Minimum; Max. - Maximum; *W* - *Shapiro*-Wilk value; *p* - p- Shapiro-Wilk value (if $P < 0.05$ does not tend to normal); 25th - 1st quartile at 25%; 50th - 2nd quartile at 50%; 75th - 3rd quartile at 75%.

Shapiro-Wilk test (Ferreira *et al.*, 2016). (Table 4), revealed that some of the indicators tend to be normal (i.e., $p > 0.05$), and others do not tend to be normal (i.e., $p < 0.05$), this allows a non-parametric approach to be adopted to conduct **hypothesis testing**. The *Kruskal-Wallis One-Way ANOVA* test was applied (Kruskal; Wallis, 1952; Chaiyo; Nokham, 2017) and the results showed (Table 5) there is a statistically significant difference in effectiveness when comparing the different indicators in the evaluations of the "Its.Redu" system.

Table 5: Hypothesis test for the effectiveness indicators of the "Its.Redu" system.

Test	**χ^2**	**df**	**P**	**ε^2**
Effectiveness	20.2	7	< 0.001	0.293

Note: χ^2 - (Chi-square), measuring the difference between groups; df - (Degrees of freedom); P - (p- *Kruskal-Wallis* value), statistical significance of the test (if $P < 0.05$ there is a significant difference between the groups);; ε^2 - (Epsilon squared), an effect size measure, representing the proportion of variance explained by the independent factor;

Significant differences are found, at a significance level of less than 0.1%, therefore there is a statistically significant difference in the indicators surveyed about the system. However, as we did not know which indicators statistically differed from each other, it was necessary to apply post-tests with multiple comparisons of *Dwass-Steel-Critchlow-Fligner* (Ladosha, 2022) (Table 6). The post-test results indicate that there is a significant statistical difference, less than *0.1%* between the indicators of "complexity and learning" when compared to "ease and use", and *4.9%* for "consistency and integration", when compared with "ease of use". Therefore, according to the results, the tutors realized that it could be effective in facilitating tutoring through the system.

Table 6: Multiple comparisons of the effectiveness indicators of the “Its.Redu” system.

Indicators		*W*	*P*
Complexity and learning	Consistency and integration	2.158	0.546
Complexity and learning	Ease of use	6.315	< 0.001
Complexity and learning	Security and trust	3.325	0.129
Complexity and learning	Utility and weight	0.819	0.978
Consistency and integration	Ease of use	3.823	0.049
Consistency and integration	Security and trust	1.884	0.671
Consistency and integration	Utility and weight	- 0.742	0.985
Ease of use	Security and trust	- 0.189	1.000
Ease of use	Utility and weight	- 3.553	0.088
Security and trust	Utility and weight	- 2.439	0.419

Note: *W* - represents the rank sum statistic between groups; *P* - is the p-value indicating statistical significance in the *Dwass-Steel-Critchlow-Fligner* analysis for multiple comparisons (if $P < 0.05$ he groups differ significantly from each other).

They also realized that with the support of the AI approach proposed by the system, it was possible to conduct activities to monitor student performance and interactions, and actively search for those with low levels of student engagement indicators, and provide personalized tutoring according to with types of difficulties. They also expressed their intention to use the “Its.Redu” in tutoring practices.

5 Considerations

In this article, we present the evaluation of a hybrid tutoring approach which both human tutors and the “Its.Redu” cooperate in mutual support to achieve effects in the teaching-learning process of students. Therefore, it is known that the challenges lie in understanding the evolution of ITS over the years. Whereas, on the one hand, human tutors often face challenges when exploring student engagement indicators, especially in the context of online learning. In some cases, the complexity of technology makes it difficult to insert systems that require specific technical skills, in addition to emerging technologies generating uncertainty and aversion to insecurity for some tutors who fear the replacement of their roles or the lack of personalization of the tutoring offered by ITS. The lack of adequate training to use ITS highlights the importance of guidance and support in integrating these systems into tutoring practice. On the other hand, the design of ITS approaches faces several challenges that involve: (i) adequately defining the domain and didactic-pedagogical purpose of the application; (ii) modeling algorithms capable of responding to learning objectives; (iii) evaluating the effectiveness to guarantee cohesive interpretations of the data and allow pedagogical actions that are mutually useful to the practices of human tutors. Furthermore, the need to ensure ethical standards requires careful validation of the multidisciplinary integration of ITS in education.

Therefore, the “Its.Redu” evaluation aims to collect human tutors’ perception towards AI. Based on the analysis, we sought to understand the effectiveness of the system's support in tutoring sessions with specific tasks aimed at tutoring activities to monitor student engagement, including recent interactions, as well as the ability to identify student interaction levels. And, during the mediation of active search, the main activity of tutors is the possibility of predicting students with low indicators of student engagement and providing suggestions based on AI algorithms for personalized tutoring according to the types of student difficulties. The evaluations revealed that while some strategies need adjustments, others can be effectively managed by algorithms. However, maintaining interpersonal interactions remains crucial of human tutoring values the performance and involvement of students in a welcoming educational movement in some cases the performance is evaded by Algorithms to cause a disturbance that

takes time to adapt to and, in the short term, can cause more disengagement than academic engagement. The limitations of "Its.Redu" pose challenges to its long-term effectiveness, as tutors may lose interest due to misalignment with dynamic instructional designs and varied educational contexts. The system's reliance on continuous data input from tutors and external platforms adds to their workload, reducing practicality. While interactive, its predefined responses limit adaptability, requiring resource-intensive model retraining to address evolving needs. Additionally, its effectiveness depends heavily on stable tutoring contexts, making it less adaptable to changes in student interactions or educational environments. Addressing these issues is essential to enhance its scalability, flexibility, and alignment with real-world educational demands.

5.1 Future works

For future work, there is the possibility of evaluating the system with professional designers specialized in the development of educational platforms. Furthermore, realizing that human tutors feel supported in a tutoring environment with the cooperation of an intelligent tutoring system. Future studies may involve looking at the adaptation of tutoring activities with strategies supported by ITS, and in this case, the questions they suggested would be related to: "*How to transfer tutoring activities to spaces in which tutors feel supported without compromising the essence of action and knowledge to be transmitted*?", and "*By using an intelligent agent that supports tutoring activities, do tutors increase the effectiveness of tutoring to promote student engagement in online learning*?". The discussions appear to be comprehensive, but ITS research, human tutor and ITS competitions, responsible and explainable approaches, and the future of strategies can define new reorientation in the evolution of ITS to support, rather than attempts to replace, human action. In this sense, in future research, we also intend to we aim to conduct new usability, attractiveness, and effectiveness tests with a larger sample size beyond the initial seven participants. Expanding the participant pool will provide more robust and generalizable insights into user experience and system performance. This will allow for a deeper understanding of potential improvements, ensuring that the AI-driven functionalities are optimized to better support tutors and students in diverse learning contexts. We have committed to conducting new tests with a larger sample of participants.

Compliance with ethical standards

It stands out: as having followed ethical and moral precepts to protect the participants involved, having no reported conflicts of interest, and having obtained the consent of the guardians involved in the research, through free and informed consent terms. As well as the commitment to current Law No. 13,709/2018 on data protection and anonymization. Furthermore, all participation was voluntary and through the signing of free and clear terms of commitment. Thus, the instruments were already calibrated in the literature and all tutors were over 18 years old.

Acknowledgements

The study participants and the FACEPE foundation for funding the research.

References

Abdallah, M., Talib, M. A., Feroz, S., Nasir, Q., Abdalla, H., & Mahfood, B. (2020). Artificial intelligence applications in solid waste management: A systematic research review. *Waste Management,* 109, (p. 231-246). https://doi.org/10.1016/j.wasman.2020.04.057. [GS Search].

Ahuja, N. J., Dutt, S., Choudhary, S. L., & Kumar, M. (2022). Intelligent Tutoring System in Education for Disabled Learners Using Human–Computer Interaction and Augmented Reality. *International Journal of Human–Computer Interaction*. https://doi.org/10.1080/10447318.2022.2124359. [GS Search].

Alhazmi, A. K., Alhammadi, F., Zain, A. A., Kaed, E., & Ahmed, B. (2023). AI's Role and Application in Education: Systematic Review. *Intelligent Sustainable Systems: Selected Papers of WorldS4 2022,* 1, (p. 1-14). https://doi.org/10.1007/978-981-19-7660-5_1. [GS Search].

Alyuz, N., Okur, E., Genc, U., Aslan, S., Tanriover, C., & Esme, A. A. (2017). An unobtrusive and multimodal approach for behavioral engagement detection of students. In: *Proceedings of the 1st ACM SIGCHI International Workshop on Multimodal Interaction for Education*. (p. 26-32). https://doi.org/10.1145/3139513.3139521. [GS Search].

Anwar, A., Haq, I. U., Mian, I. A., Shah, F., Alroobaea, R., Hussain, S., ... & Umar, F. (2022). Applying real-time dynamic scaffolding techniques during tutoring sessions using intelligent tutoring systems. *Mobile Information Systems*. https://doi.org/10.1155/2022/6006467. [GS Search].

Ashwin, T. S., & Guddeti, R. M. R. (2020). Affective database for e-learning and classroom environments using Indian students' faces, hand gestures and body postures. *Future Generation Computer Systems*, v. 108, (p. 334-348). https://doi.org/10.1016/j.future.2020.02.075. [GS Search].

Aslan, S., Alyuz, N., Okur, E., Mete, S. E., Oktay, E., & Esme, A. A. (2018). Effect of emotion-aware interventions on students' behavioral and emotional states. *Educational Technology Research and Development*, 66(6), (p. 1399-14130). https://doi.org/10.1007/s11423-018-9589-7. [GS Search].

Baothman, F. A. (2021). An intelligent big data management system using haar algorithm-based Nao agent multisensory communication. *Wireless Communications and Mobile Computing*. https://doi.org/10.1155/2021/9977751. [GS Search].

Chaiyo, Y.; & Nokham, R. (2017). The effect of Kahoot, Quizizz and Google Forms on the student's perception in the classrooms response system. In *2017 International Conference on Digital Arts, Media and Technology (ICDAMT)*. IEEE. (p. 178-182). https://doi.org/10.1109/ICDAMT.2017.7904957. [GS Search].

Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two Decades of Artificial Intelligence in Education. *Educational Technology & Society*, 25(1), (p. 28-47). Available at: https://www.jstor.org/stable/48647028. [GS Search].

Chen, H., Park, H. W., & Breazeal, C. (2020). Teaching and learning with children: Impact of reciprocal peer learning with a social robot on children's learning and emotive engagement. *Computers & Education*, v. 150. https://doi.org/10.1016/j.compedu.2020.103836. [GS Search].

Chrysafiadi, K., Virvou, M., Tsihrintzis, G. A., & Hatzilygeroudis, I. (2022). Evaluating the user's experience, adaptivity and learning outcomes of a fuzzy-based intelligent tutoring system for computer programming for academic students in Greece. *Education and*

Information Technologies, (p. 1-31). https://doi.org/10.1007/s10639-022-11444-3. [GS Search].

Conati, C., Barral, O., Putnam, V., & Rieger, L. (2021). Toward personalized XAI: A case study in intelligent tutoring systems. *Artificial intelligence*, 298. https://doi.org/10.1016/j.artint.2021.103503. [GS Search].

De Almeida, G. M., Gomes, A. S., de Almeida, J. C., Nogueira, T. J. D. D., Lima, R. M. F., de Melo, T. S., & de Mello, L. A. (2020). Evaluating Entrepreneurial Perceptions on Blended Learning. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. https://doi.org/10.1109/SMC42975.2020.9282851 [GS Search].

D'Mello, S. K., & Graesser, A. (2012). Language and discourse are powerful signals of student emotions during tutoring. *IEEE Transactions on Learning Technologies*, 5(4), (p. 304-317). https://doi.org/10.1109/TLT.2012.10. [GS Search].

Echeverria, V., Yang, K., Lawrence, L., Rummel, N., & Aleven, V. (2023). Designing Hybrid Human-AI Orchestration Tools for Individual and Collaborative Activities: A Technology Probe Study. *IEEE Transactions on Learning Technologies*. https://doi.org/10.1109/TLT.2023.3248155. [GS Search].

Ferreira, H. N. M., Brant-Ribeiro, T., Araújo, R. D., Dorça, F. A., & Cattelan, R. G. (2016). An automatic and dynamic student modeling approach for adaptive and intelligent educational systems using ontologies and bayesian networks. In *2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI)* IEEE. (p. 738-745). https://doi.org/10.1109/ICTAI.2016.0116. [GS Search].

Gao, M., Kortum, P., & Oswald, F. L. (2020). Multi-language toolkit for the system usability scale. *International Journal of Human–Computer Interaction*, 36(20), (p. 1883-1901). https://doi.org/10.1080/10447318.2020.1801173. [GS Search].

Gilbert, S. & Dorneich, M. (2018). Applying Human-Agent Team Concepts to the Design of Intelligent Team Tutoring Systems. In: *Intelligent Tutoring Systems: 14th International Conference, ITS 2018, Montreal, QC, Canada, June 11–15, 2018, Proceedings*. Springer. http://doi.org/10.1007/978-3-319-91464-0. [GS Search].

Kar, A. K., Choudhary, S. K., & Singh, V. K. (2022). How can artificial intelligence impact sustainability: A systematic literature review. *Journal of Cleaner Production.* https://doi.org/10.1016/j.jclepro.2022.134120. [GS Search].

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260), (p. 583-621). https://doi.org/10.1080/01621459.1952.10483441. [GS Search].

Kurita, T. (2019). Principal component analysis (PCA). *Computer Vision: A Reference Guide*, (p. 1-4). https://doi.org/10.1007/978-3-030-03243-2_649-1. [GS Search].

Lacerda, D. P., Dresch, A., Proença, A., & Antunes Júnior, J. A. V. (2013). Design Science Research: método de pesquisa para a engenharia de produção. *Gestão & produção,* 20(4), (p. 741-761). https://doi.org/10.1590/S0104-530X2013005000014. [GS Search].

Ladosha, O. M. (2022). Dynamics of Student Performance in a Foreign Language from the Perspective of the Transition to Distance Learning (statistical analysis). In: *2022 VI International Conference on Information Technologies in Engineering Education (Inforino)*. IEEE. (p. 1-5). https://doi.org/10.1109/Inforino53888.2022.9782984. [GS Search].

Latham, A. (2022). Conversational Intelligent Tutoring Systems: The State of the Art. *Women in Computational Intelligence: Key Advances and Perspectives on Emerging Topics*. (p. 77-101). https://doi.org/10.1007/978-3-030-79092-9_4. [GS Search].

Li, S., Lajoie, S. P., Zheng, J., Wu, H., & Cheng, H. (2021). Automated detection of cognitive engagement to inform the art of staying engaged in problem-solving. *Computers & Education*, v. 163, (p. 104-114). https://doi.org/10.1016/j.compedu.2020.104114. [GS Search].

Liu, H. L., Wang, T. H., Lin, H. C. K., Lai, C. F., & Huang, Y. M. (2022). The Influence of Affective Feedback Adaptive Learning System on Learning Engagement and Self-Directed Learning. *Frontiers in Psychology*, 13(1). https://doi.org/10.3389/fpsyg.2022.858411. [GS Search].

Lourenço, D. F., Carmona, E. V., & Lopes, M. H. B de M. (2022). Tradução e adaptação transcultural da System Usability Scale para o português do Brasil. *Aquichan*, 22(2), (p. 8). https://doi.org/10.5294/aqui.2022.22.2.8. [GS Search].

Maldaner, N., Pozzebon, E., & Dos Santos, T. N. (2022). Proposta de implementação da Computação Afetiva no Sistema Tutor Inteligente MAZK: conciliando emoções com o processo de aprendizagem. *RENOTE*, 20(2). https://doi.org/10.22456/1679-1916.129153. [GS Search].

Martins, A. I., Rosa, A. F., Queirós, A., Silva, A., & Rocha, N. P. (2015). European Portuguese validation of the system usability scale (SUS). *Procedia computer science*, v. 67, (p. 293-300). https://doi.org/10.1016/j.procs.2015.09.273. [GS Search].

Mithas, S., Chen, Z. L., Saldanha, T. J., & De Oliveira Silveira, A. (2022). How will artificial intelligence and Industry 4.0 emerging technologies transform operations management?. *Production and Operations Management*, 31(12), (p. 4475-4487). https://doi.org/10.1111/poms.13864. [GS Search].

Muñoz, K., Mc Kevitt, P., Lunney, T., Noguez, J., & Neri, L. (2011). An emotional student model for game-play adaptation. *Entertainment Computing*, 2(2), (p. 133-141). https://doi.org/10.1016/j.entcom.2010.12.006. [GS Search].

Nielsen, J. (2000). *Why you only need to test with 5 users*. Available at: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users. [GS Search].

Ogunyemi, A. A., Quaicoe, J. S., & Bauters, M. (2022). Indicators for enhancing learners' engagement in massive open online courses: A systematic review. *Computers and Education Open*. https://doi.org/10.1016/j.caeo.2022.100088. [GS Search].

Okur, E., Alyuz, N., Aslan, S., Genc, U., Tanriover, C., & Arslan Esme, A. (2017). Behavioral engagement detection of students in the wild. In: *International Conference on Artificial*

Intelligence in Education. Springer, Cham. (p. 250-261). https://doi.org/10.1007/978-3-319-61425-0_21. [GS Search].

Pereira, A. J., Gomes, A. S., Primo, T. T., da Silva, R. M. A., Rodrigues, R. L., de Campos Filho, A. S., ... & de Melo Júnior, R. P. (2021). Identificação e caracterização de níveis de interação no ensino remoto de emergência na Educação Básica. In: *Anais do XXXII Simpósio Brasileiro de Informática na Educação. SBC, 2021*. (p. 145-156). https://doi.org/10.5753/sbie.2021.218498. [GS Search].

Pereira, A. J., Gomes, A. S., & Primo, T. T. (2022). Design de Sistema de Recomendação Educacional: abordagens com Mágico de Oz. In: *Anais do XXXIII Simpósio Brasileiro de Informática na Educação. SBC, 2022*. (p. 1184-1195). https://doi.org/10.5753/sbie.2022.225760. [GS Search].

Pereira, A. J., Gomes, A. S., Primo, T. T., Rodrigues, R. L., Júnior, R. P. M., & Moreira, F. (2023). Learning Mediated by Social Network for Education in K-12: Levels of Interaction, Strategies, and Difficulties. *Education Sciences*, 13(2), (p. 100). https://doi.org/10.3390/educsci13020100. [GS Search].

Pereira, A. J., Gomes, A. S., & Primo, T. T. (2023a). Desvendando a Resistência e as Estratégias dos Tutores Humanos para Engajar Estudantes no Aprendizado Online. *Cadernos de Educação Tecnologia e Sociedade*, 16(3), (p. 551-566). https://doi.org/10.14571/brajets.v16.n3.551-566. [GS Search].

Pereira, A. J.; Gomes, A. S.; & Primo, T. (2023b). Uma Abordagem de Sistema de Tutoria Inteligente para Cooperação com a Atuação de Tutores Humanos. *RENOTE*, Porto Alegre, 21(2), (p. 208–219). https://doi.org/10.22456/1679-1916.137742. [GS Search].

Pereira, A. J., Gomes, A. S., Primo, T. T., Queiros, L. M., & Moreira, F. (2024). Anticipating Tutoring Demands Based on Students Difficulties in Online Learning. *Lecture Notes in Computer Science*. 1ed.: Springer Nature Switzerland, (p. 321-332). http://dx.doi.org/10.1007/978-3-031-61691-4_21. [GS Search].

Pereira, A. J., De Almeida, G. M., Gomes, A. S., & Primo, T. T. (2024a) Identificação de Estilos de Tutoria no Aprendizado Online. In: *Simpósio Brasileiro de Informática na Educação (SBIE)*, 35, 2024, Rio de Janeiro/RJ. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação. (p. 1336-1348). https://doi.org/10.5753/sbie.2024.242537. [GS Search].

Pereira, A. J., Gomes, A. S., & Primo, T. (2024). Classificação de Interações com Indicadores de Engajamento dos Estudantes no Aprendizado Online. In: *Simpósio Brasileiro de Informática na Educação (SBIE)*, 35, 2024, Rio de Janeiro/RJ. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação. (p. 1574-1586). https://doi.org/10.5753/sbie.2024.242141. [GS Search].

Pereira, A. J., Gomes, A. S., & Primo, T. (2024a). Funcionalidades Emergentes para Tutorias em Ambientes Virtuais de Aprendizagem. *Revista iberoamericana de tecnologia en educación y educación en tecnología* (en línea), v. 38, (p. 64-71). https://doi.org/10.24215/18509959.38.e6. [GS Search].

Pereira, A. J.; Gomes, A. S.; & Primo, T. (2025). Paths of Socially Shared Regulation of Learning Students in Online Mediated Activities. *RENOTE*, Porto Alegre, 22(3), (p. 11–20, 2025). http://dx.doi.org/10.22456/1679-1916.144965. [GS Search].

Pham, P., & Wang, J. (2018). Predicting learners' emotions in mobile MOOC learning via a multimodal intelligent tutor. In: *International Conference on Intelligent Tutoring Systems*. Springer, Cham. (p. 150-159). https://doi.org/10.1007/978-3-319-91464-0_15. [GS Search].

Phillips, A., Pane, J. F., Reumann-Moore, R., & Shenbanjo, O. (2020). Implementing an adaptive intelligent tutoring system as an instructional supplement. *Educational Technology Research and Development*, 68(3), (p. 1409-1437). https://doi.org/10.1007/s11423-020-09745-w. [GS Search].

Pimentel, M., Filippo, D., & Dos Santos, T. M. (2020). Design Science Research: pesquisa científica atrelada ao design de artefatos. *RE@ D-Revista de Educação a Distância e eLearning*, 3(1), (p. 37-61). https://doi.org/10.34627/vol3iss1pp37-61. [GS Search].

Pink, S., Gomes, A., Zilse, R., Lucena, R., Pinto, J., Porto, A., ... & De Oliveira, M. D. (2018). Automated and connected? Smartphones and automobility through the global south. *Applied Mobilities*. https://doi.org/10.1080/23800127.2018.1505263. [GS Search].

Ponticorvo, M., Ferrara, F., Di Fuccio, R., Di Ferdinando, A., & Miglino, O. (2017). SNIFF: A game-based assessment and training tool for the sense of smell. In: *Methodologies and Intelligent Systems for Technology Enhanced Learning: 7th International Conference. Springer International Publishin*. (p. 126-133). https://doi.org/10.1007/978-3-319-60819-8_15. [GS Search].

Razak, T. R., Ismail, M. H., Fauzi, S. S. M., Gining, R. A. J., & Maskat, R. (2021). A framework to shape the recommender system features based on participatory design and artificial intelligence approaches. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 10(3), (p. 727). https://doi.org/10.11591/ijai.v10.i3.pp727-734. [GS Search].

Reis, S. C., Gomes, A. F., & De Souza, R. S. (2014). Explorando a Rede Social Educacional no ensino de línguas: possibilidades, gêneros e multiletramentos. *RENOTE*, 12(1). https://doi.org/10.22456/1679-1916.50280. [GS Search].

Ringnér, M. (2008). What is principal component analysis?. *Nature biotechnology*, 26(3), (p. 303-304). https://doi.org/10.1038/nbt0308-303. [GS Search].

Rogers, Y., Sharp, H., & Preece, J. *Design de Interação*. Bookman ed. [s.l: s.n.]. [GS Search].

Ruan, S., He, J., Ying, R., Burkle, J., Hakim, D., Wang, A., ... & Landay, J. A. (2020). Supporting children's math learning with feedback-augmented narrative technology.?. In: *Proceedings of the Interaction Design and Children Conference*. (p. 567-580). https://doi.org/10.1145/3392063.3394400. [GS Search].

Ruiz, N., Yu, H., Allessio, D. A., Jalal, M., Joshi, A., Murray, T., ... & Betke, M. (2022). ATL-BP: a student engagement dataset and model for affect transfer learning for behavior prediction. *IEEE Transactions on Biometrics, Behavior, and Identity Science*. https://doi.org/10.1109/TBIOM.2022.3210479. [GS Search].

Sadiq, R. B., Safie, N., Abd Rahman, A. H., & Goudarzi, S. (2021). Artificial intelligence maturity model: a systematic literature review. *PeerJ Computer Science*, 7. https://doi.org/10.7717/peerj-cs.661. [GS Search].

Srinivasa, K. G., Kurni, M., & Saritha, K. (2022). Harnessing the Power of AI to Education. In *Learning, teaching, and assessment methods for contemporary learners: pedagogy for the digital generation* (p. 311-342). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-19-6734-4_13. [GS Search].

St-Hilaire, F., Vu, D. D., Frau, A., Burns, N., Faraji, F., Potochny, J., ... & Kochmar, E. (2022). A New era: Intelligent tutoring systems will transform online learning for millions. *arXiv preprint arXiv:2203.03724.* https://doi.org/10.48550/arXiv.2203.03724. [GS Search].

Wamba-Taguimdje, S. L., Wamba, S. F., Kamdjoug, J. R. K., & Wanko, C. E. T. (2020). Influence of artificial intelligence (AI) on firm performance: the business value of AI-based transformation projects. *Business Process Management Journal*, 26(7), (p. 1893-1924). https://doi.org/10.1108/BPMJ-10-2019-0411. [GS Search].

Wang, W., & Siau, K. (2019). Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda. *Journal of Database Management (JDM),* 30(1), (p. 61-79). http://doi.org/10.4018/JDM.2019010104. [GS Search].

Chiu, T. K., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2022). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*. https://doi.org/10.1016/j.caeai.2022.100118. [GS Search].

Xiao, X., & Wang, J. (2016). Context and cognitive state triggered interventions for mobile MOOC learning. In: *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. (p. 378-385). https://doi.org/10.1145/2993148.2993177. [GS Search].

Zafari, M., Bazargani, J. S., Sadeghi-Niaraki, A., & Choi, S. M. (2022). Artificial intelligence applications in K-12 education: A systematic literature review. *IEEE Access*. https://doi.org/10.1109/ACCESS.2022.3179356. [GS Search].