

AN EXPERIENCE ON THE USE OF CHATGPT AS AN EDUCATIONAL TOOL  
FOR CREATING PRIMARY EFL CLASSROOM MATERIALS AND  
ACTIVITIES: PRE-SERVICE TEACHERS' PERCEPTIONS AND EXPERTS'  
OPINIONS

---

AMAIA AGUIRREGOITIA MARTÍNEZ

*University of the Basque Country EHU-UPV*

PAULO KORTAZAR BILLELABEITIA

*University of the Basque Country EHU-UPV*

RUTH MILLA

*University of the Basque Country EHU-UPV*

## 1. INTRODUCTION

The present study was conducted in a Teaching Faculty in Spain to answer the following question: Can ChatGPT assist teachers in creating pedagogical proposals and produce practical educational activities/materials? The experiment involved a total of 121 third-grade students enrolled in a foreign language teaching (FLT) course, who were required to create communicative (CLT) activities and materials for a flipped classroom (FC) lesson using ChatGPT as an assistant and then record the commands fed. Later, they answered a questionnaire (1-10 scale) on the usefulness of the tool and the modifications they made to the proposal generated by ChatGPT. Additionally, students evaluated whether the tool could substitute for the teacher's job and whether they would use it in their future teaching. The questionnaire also included open questions on the pros and cons of using the tool along with some ethical issues regarding authorship. Statistical tests and qualitative analyses were performed on the data. Finally, the students' creations' appropriateness and practicality were measured by

an experts' committee. Results suggest that ChatGPT may be a useful tool for material design and lesson planning, although potential problems regarding authorship and pedagogical issues were identified both by the participant students and the expert evaluators, as we will explain later on.

In what follows, we will review existing literature on the use of artificial intelligence and its benefits for education, examples of experiences with ChatGPT and issues related to plagiarism. We will then present the objectives and methodology followed for the present study. Besides, results will be described and discussed, as well as some conclusions derived from the findings.

## 2. AI IN EDUCATION

The term artificial intelligence (AI) was first coined by John MacCarthy at a conference in 1956 (Tahiru, 2021) to refer to the development of machines that mimic human behaviour. In the latest decades, science has evolved by using machine learning and algorithms in multiple devices and applications that pervade everyday life. One of the realms where AI has become commonplace is education, with a variety of programmes designed to aid teachers, learners and administrations in their tasks. Artificial intelligence in education (AIEd) has been massively spread in the last decade, especially with the pandemic, when technology was more than assisting, an essential means through which to deliver content to students all over the world. The application of AIEd derives from the work of the psychologists Pressey (1950) and Skinner (1958) and relies on learning theories such as Vygotsky's zone of proximal development (1978) and Gagné's instructionalism (1985).

### 2.1. USES

Holmes et al (2019) explain that AIEd is growing exponentially and so is research, but still there is very limited knowledge of the potential in the classroom and how to use it most effectively. The authors mention different uses of AIEd, which

includes everything from AI-driven, step-by-step personalized instructional and dialogue systems, through AI-supported exploratory learning, the analysis of student writing, intelligent agents in game-based environments, and student-support chatbots, to AI-facilitated student/tutor matching that puts students firmly in control of their own learning. It also includes students interacting one-to-one with computers, whole-school approaches, students using mobile phones outside the classroom, and much more besides.

(Holmes et al 2019, p.11)

Ouyang and Jiao (2021) also mention the lack of results on the effect of the use of AI and technologies on learning and instruction. In their review of previous research on AIED, they distinguish three different paradigms with the learner as a recipient, the learner as a collaborator and the learner as a leader. Examples of implementation of these three paradigms would be intelligent tutoring systems (Tahiru, 2021), dialogue-based tutoring systems, and personalised/adaptive learning, respectively.

Machine learning has been applied to foreign language learning applications (Vincent-Lancrin & van der Vlies, 2020), with machines that provide learners with feedback on pronunciation, for example. Another use of AIED is for assessment (Tahiru, 2021), such as game-based assessment using virtual or augmented reality, which is especially useful for formative assessment but also for the summative type (Vincent-Lancrin & van der Vlies, 2020).

## 2.2. BENEFITS

Since AIED stems from pedagogical and theoretical foundations about the learning process, its application results in an evolution of the teacher's and learner's role, leading to promote learner-centred and personalised processes and fostering learners' autonomous learning (Chen et al, 2020; Ouyang & Liao, 2021). AIED helps learners advance "at their own pace and provides teachers with suggestions on how to help them" (Vincent-Lancrin & van der Vlies 2020, p.7). This individualised learning process, in turn, proves beneficial for supporting students with special needs (Vincent-Lancrin & van der

Vlies. 2020). Additionally, critical thinking skills and creativity are further developed by means of technology (Chen et al, 2020; Vincent-Lancrin & van der Vlies, 2020).

### 2.3. CHATGPT: WHAT IT IS, BRIEF HISTORY, USE AND POSSIBILITIES FOR EDUCATION.

Atlas (2023, p.2) defines ChatGPT (Generative Pre-trained Transformer) as “a powerful, cutting-edge language model developed by OpenAI that uses artificial intelligence to generate text that is similar to human writing. ChatGPT derives from GPT-3 and its release to the public on November 30, 2022 reflected a significant step forward in language models technology.” The emergence of ChatGPT (and its use in educational contexts) has been compared to the appearance of calculators, a milestone for maths and other science-related subjects, which was really controversial at the beginning, but later teachers adapted the way they assess learning and the calculators’ real practicality and advantages were seen.

AI in the form of ChatGPT can assist teachers in their material creation process and lesson planning, as well as being useful in grading, assessing students’ work and designing quizzes and other evaluation tools. In this line, Perlman wrote a paper together with ChatGPT where, in spite of the inadequacies shown by the tool, a wide range of possibilities can be seen for different realms of society (ChatGPT & Perlman, 2022).

However, in order to obtain the maximum effectiveness from the tool, educators should be made aware of its limitations and potentialities and trained on its use (Atlas, 2023). Similarly, some authors point out that the limitations and threats posed by the use of ChatGPT by students can be fixed by changing the way educators evaluate and the types of questions we ask (Baidoo-Anu & Owusu Ansah, 2023, Frye, 2022). Additionally, Bishop (2023) concluded that the tool is not capable of creativity or critical thinking, so this is the skill educators should be aiming to develop in their students to prevent them from overusing or ill-using the tool (Codina, 2022). Other authors also mention the lack

of logical reasoning (Llorens-Largo & ChatGPT, 2022) and the occasional errors (Bowman, 2022; Pearl, 2002) in the responses provided by ChatGPT.

## 2.4. PREVIOUS EXPERIENCES

### 2.4.1. Experiences with AI in higher education

In tertiary education, one of the problems feared by educators is the plagiarism and potential overuse of ChatGPT in the students' assignments. Some ideas on how to avoid these problems have been mentioned above, which consist of taking advantage of the tool's deficits when designing the tasks. Pickell and Doak (2023) also propose asking the students to include citations, personal experiences and critical discussion in their writings, which ChatGPT cannot do, as well as insert face-to-face subtasks in the middle of projects, to follow up on students' progress and make sure they are not relying on this type of tools exclusively. Also regarding higher education, Rudolph et al (2023) thoroughly analyse ChatGPT's strengths and weaknesses and review existing literature on the tool, which, according to the authors, consists basically of opinion papers. The authors highlight some advantages of using ChatGPT, such as the fact that students learn through experimentation, that teamwork and collaboration are promoted, and the flexible, quick and cost-effective nature of the tool (Rudolph et al, 2023). As mentioned above, the authors consider ChatGPT a valuable tool for students to "help spark the creative process (McMurtrie, 2023, cited in Rudolph et al 2023, p.14)" but students should be trained to use it sensibly and combine its products with critical thinking and creativity.

### 2.4.2. Experiences with AI in primary education

As mentioned above, the majority of the literature on AIED is devoted to expressing the authors' views and predictions on the different tools. There are a few more practical publications that offer educators ideas and tips on how to use them most effectively but, as we have seen in the previous sections, they are limited to higher education. However,

lower stages of education, such as primary, have remained rather overlooked by researchers, preventing educators at this level from taking advantage of AI's possibilities for their classes.

An interesting paper on the application of AIED was Srinivasan and Murthy (2021) in K-12 in India, probably the only large-scale study on the use of AI for developing children's literacy in English. The implementation of the technology in more than 5000 schools led to 20-40% gains in learning outcomes (rising to 50-60% in the self-administered assessments).

Additionally, Williamson and Eynon (2020) call for more ethnographic case-study work on AI in different educational contexts to enrich existing research on the potentialities of AIED, which tend to be rather general and do not illustrate the reality of AIED in the actual classrooms. Besides, we consider that studies oriented to primary school educators, like the one presented in this chapter, may be helpful to fill the research gap on the use of AI for syllabus design and material creation.

### 3. THE QUESTION OF PLAGIARISM

#### 3.1. PLAGIARISM IN HIGHER EDUCATION

Studies in the last decade have pointed to an increase in the proportion of university students who commit plagiarism (Howard, Enrich & Walton, 2014). Some sources indicate that at some point in their careers as high as three quarters of students have engaged in practices involving academic dishonesty (Brimble & Stevenson-Clarke, 2005). The studies in plagiarism have branched into different areas, of which we shall list three that are related to our setting and participants.

First, there is a trend in studies of plagiarism that push for an understanding of plagiarism and derived practices tied to the cultural mindset. As Hu and Lei claim, the understanding of plagiarism varies among disciplines while also being “a complex, multidimensional phenomenon mediated by a variety of factors and embedded deeply in cultural, social, historical, ideological, and epistemological conditions”

(2005, p. 234). Second, the research in the area has also treated the self-reported and applied observations that students hold on plagiarism as separate elements (Pittam et al, 2009; Risquez et al, 2009). As Risquez et al declare

it is possible that students understand what plagiarism is in concept, declare to hold supporting ethical views and expectations of punishment and even claim to refrain from committing the offence, but they may not necessarily recognise it in practice as a breach of academic guidelines to be avoided.

(2009, p. 35).

That is, even if students believe that plagiarism constitutes an ethical breach, this position might be contradicted by the fact that in practice some students may engage in unethical behaviour. Third, plagiarism in L2 is identified as a staple in undergraduates' writings due to the fact that writers tend to use sources without acknowledging them (Pecorari, 2015) or due to their lack of writing proficiency they engage in a practice called "patchwriting", by which extracts from other sources are used to weave a text (Howard, 1995; Pecorari, 2015).

### 3.2. CHATGPT AND PLAGIARISM

The worries that students may overuse ChatGPT (Codina, 2022) and the fact that the generative capacity of the AI may lead to changes in academic assessment (Pickell & Doak, 2023) can be understood in the abovementioned context of plagiarism as a serious issue in higher education. According to Cotton et al, the challenges that ChatGPT poses for assessment are that the AI can generate text, which is in turn difficult to discern from that written by a student (2023). This could lead to the widespread use of the tool among students (Dehouche, 2021). Reactions to its unethical use have been varied, including the prediction of the end of academic essays (Stokel-Walker, 2023).

In contrast, many researchers have taken a more critical approach when analysing the potential of this AI to write an essay. Osmanovic-Thunström and Steingrimsson (2023) relativise its potential, while others lean towards a change in assessment criteria,

tasks and strategies in order to adapt to the new context (Cotton et al, 2023).

The applied research in the field has concentrated on the ability that AI-generated writings have to pass detector tests (Anderson et al, 2023; Gao et al, 2023). As per Gao et al.'s findings, human reviewers were able to identify 68% of AI-generated abstracts, whereas the GPT-2 Output Detector was capable of identifying 99.98% of fake abstracts (2023).

#### 4. OBJECTIVES

The general objective of this paper is to answer the following research question: Can ChatGPT assist teachers in creating pedagogical proposals and producing practical educational activities/materials?

Three specific objectives were defined for that purpose:

1. To explore the potential of ChatGPT for activity design and material creation in FLT.
2. To apply experts' judgement to evaluate the design of activities and material according to CLT.
3. To analyse students' perception of authorship and plagiarism issues, when using ChatGPT.

The next section details the methodology used to achieve these objectives.

#### 5. METHODOLOGY

##### 5.1. CONTEXT AND PARTICIPANTS

Four cohorts were used to analyse students' perceptions of ChatGPT and the real use and possibilities of this tool. The participants were 121 third-grade students enrolled in the degree of Primary School Teaching at the Faculty of Education in Bilbao. The study involved 87 women, 33 men and 1 non-binary person aged 20 to 26. The FLT courses involved were taught in the second term of 2022-23. During this term, more specifically in one of the courses and for three months, the



students had been taught the basic didactics of English as a foreign language and they have been in contact with the principles of CLT and FC. The course has an emphasis on the practical part of analysing current materials and designing communicative activities for primary education students. The second task proposed in the course involves the creation of a lesson using the FC technique where students have to prepare a video to teach some content of Primary Education (about Social Science, Science, Maths...). Additionally, students had to design some activities both to control that potential primary students have watched the video at home and also to cooperatively create something in the classroom by applying their knowledge. This second task of the course is the one that students completed with the assistance of ChatGPT after some initial instruction on the use of the chatbox.

## 5.2. DATA GATHERING TOOLS AND DATA ANALYSIS TOOLS

The source of data that informed the study is an anonymous survey including questions about the effective use of ChatGPT in each of the different sub-tasks: write the introduction to the topic, present the specific content, prepare texts, the layer of control and the communicative activities, and adapt the language for primary education students. Participants kept track of the number of iterations and questions typed to ChatGPT to obtain a valid answer they considered appropriate for their group of students and topic and they included this information in their final report. Additionally, they were asked to assess from 1 to 10 the usefulness of the tool for each of the subtasks and the extent to which modifications had been made to the original proposal made by the tool. Students were also asked to evaluate the degree to which the tool can substitute the teacher's job and if they would use it when teaching in the future. Finally, some open questions were designed to invite participants to reflect on the most beneficial aspects and potential problems of using the tool along with some ethical issues regarding the authorship of the materials.

To measure internal consistency, Cronbach's tests have been carried out. The result of the test using the data obtained from questions 1 to 6 (usability of the tool) of the questionnaire is  $\alpha = 0.78$ , which confirms

that the data are acceptably consistent. Additionally, that test has also been carried out on the data from questions 21, 22 and 23 (related to authorship) and the value for the alpha of Cronbach is  $\alpha= 0.64$ , which indicates a more questionable consistency.

Face validity has been tested using the assessments by three experts in the field. Students' numerical answers have been analysed to obtain mean values and standard deviations.

A qualitative analysis was carried out on the answers to the open questions to identify and interpret patterns in the information reported about the strengths, weaknesses, potential problems the students had recognized, and some additional aspects related to authorship and plagiarism. The patterns found in these answers are described in the results section.

### 5.3. THE INTERVENTION

Before the intervention, students were instructed on the basic theoretical and practical aspects of CLT and the design of activities aligned with these principles. Then, two theoretical sessions on the principles of the FC technique took place followed by a session to introduce ChatGPT along with its possibilities. Students had the opportunity to observe the answers provided by the tool to specific commands aimed at text creation, design of activities and adaptation of the language used to deal with the inquiry made. Later, two practical sessions and six hours of group work outside the classroom were planned. Finally, the students submitted their assignments and an online survey about the experience was carried out.

After the intervention, the experts assessed the proposals independently, assigning an individual mark to each of the criteria and adding their comments about the deficiencies and strengths found in the proposals.

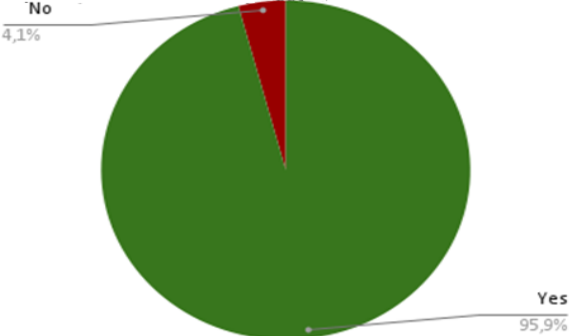
## 6. RESULTS

### 6.1. RESULTS: CHATGPT AS AN EDUCATIONAL RESOURCE.

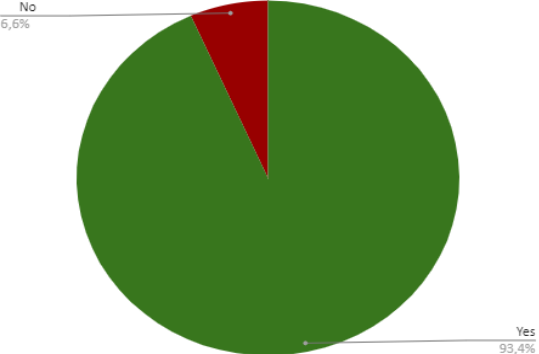
The next lines detail the results obtained using the online questionnaire.

Regarding the usefulness of the tool, students consider it convenient and claim that they would use it to create materials, explanations and activities (figures 1 and 2), and also in their assignments at the university.

**FIGURE 1.** Percentage of students that would use ChatGPT to produce materials and explanations



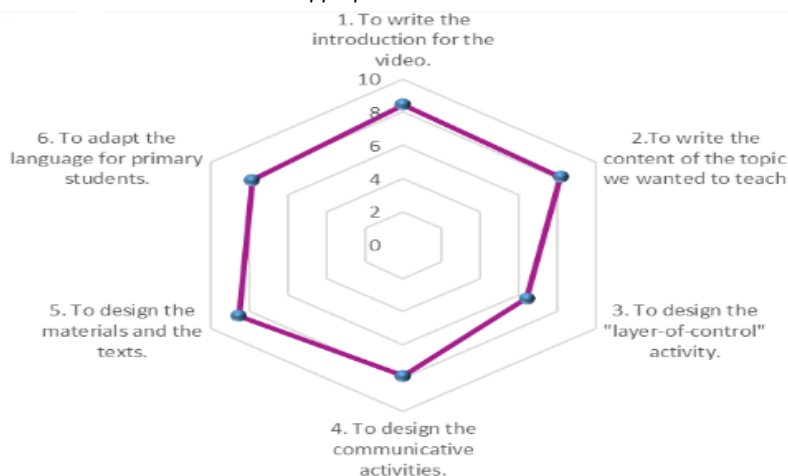
**FIGURE 2.** Percentage of students that would use ChatGPT to create activities for the classroom



Additionally, the assessment of the tool for the different subtasks involved in the assignment reveals that students consider it practical

for every subtask. Figure 3 shows the mean value is between 7.8 and 8.45 for every subtask, except for one. In this exceptional case (the task about designing a control task to check if students have watched and understood the video), the mean value is 6.42 (see also Table 1 below for descriptive statistics). However, the reason behind it is that students had not understood what they were expected to do at this particular point.

**FIGURE 3.** Assessment of the appropriateness of the tool for the different subtasks



**TABLE 1.** Statistics of the degree of conformity expressed with the statement "The tool has been useful for each of the subtasks (0-Lowest 10- Highest)"

	ST1	ST2	ST3	ST4	ST5	ST6
Mean	8.45	8,17	6.42	7.87	8.53	7.84
Deviation	1.62	1.79	2.77	1.86	1.39	1.93
Median	10	10	8	10	10	10
Mode	10	10	10	10	10	10

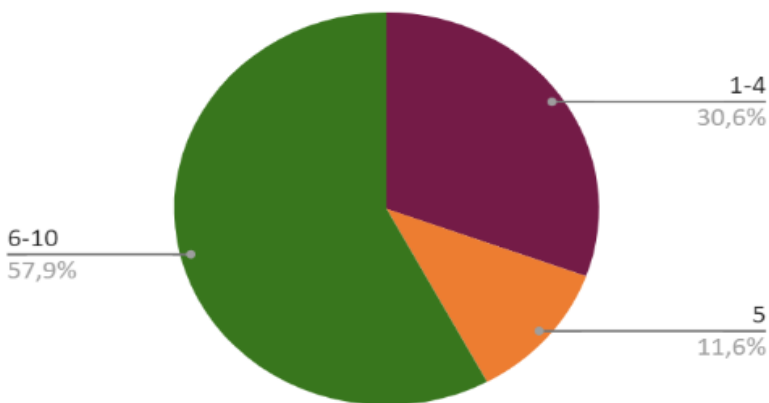
Note: n=121

ST1. To write the introduction for the video; ST2. To write the content of the topic we wanted to teach; ST3. To design the "layer-of-control" activity; ST4. To design the communicative activities; ST5. To design the materials and the texts; ST6. To adapt the language for primary students.

Additionally, most students admit that they would use it in future university assessment tasks: 84 students answered affirmatively, 7 responded negatively and 21 said they were uncertain.

The test also asked the students about their level of agreement with some statements on a scale from 1 to 10. Figure 4 shows that 69.5 % of the students identify that the proposals made by the tool require adjustments.

**FIGURE 4.** Degree of agreement with the statement “Modifications had to be made to the proposal of the tool”



When asked about the number of iterations or questions made to the tool to obtain the desired result, most students admitted having used no more than three questions. Table 2 shows that they tended to accept the initial proposals made by the tool for every subtask in most cases.

**TABLE 2.** Percentage of students that used 3 or fewer iterations to solve the subtask

Subtask	Percentage of students that used 3 or fewer iterations
7.To achieve an appropriate video introduction	60.10 %

Subtask	Percentage of students that used 3 or fewer iterations
8.To obtain the theory on the content you wanted to teach	57 %
9.To design an appropriate layer of control activity	72.40 %
10.To design at least three appropriate communicative activities	62.60 %
11.To create appropriate material and texts	49.60 %
12. To adapt the language to the level of primary students	66.70 %

As Table 2 shows, the commands fed to the AI were not above 3 in critical areas such as in the design of the layer of control activity (in the case of 72.40% of students), the design of communicative activities (in the case of 62.60% of students), design of appropriate materials and texts (in the case of 49.60% of students) and when adapting the output that the chat produced to the language level of primary students (in the case of 66.70% of students). Other areas that were not directly related to the design of activities, but were crucial to the creation of the FC video, corroborate this tendency showing less than three iterations when aiming to obtain a video introduction (for 60.10% of students) and content to teach on the video (for 57% of students).

Regarding the modifications made, the set of data is spread out. The calculated mean value is 5.71 for this item with a standard deviation of 2.24 indicating that the data is widely scattered.

As to the open questions about the strengths and weaknesses of the tool, students highlighted that it generates new ideas, offers reliable information, is valid to offer explanations on specific concepts/topics, and creates materials and activities that can be used in the classroom as they are proposed. They also assessed positively the possibilities it offers to adapt the texts to the level and age, along with its usability.

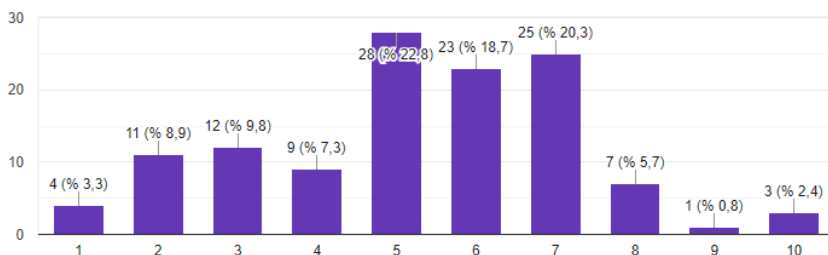
The open questions allowed the identification of some of the most common concerns. They mentioned that the proposals may not be

appropriately adapted to the characteristics of the classroom, special needs, and objectives of the particular group of students and that it may be difficult for the tool to adequately adapt the language and vocabulary to the level required. They also emphasised that information provided by the tool may be unreliable and that it is not easy to know about its source. Regarding the consequences of relying on the tool, they predict that teachers may lose their autonomy and critical thinking skills if they count too much on the tool and become dependent on its use. They stated that teachers may stop producing their own ideas and cultivating originality, creativity and imagination. They also noted that the importance of the teacher may be ignored or replaced as an effect. However, 73.7% of the respondents disagree with the idea of the tool being a substitute for the teacher.

## 6.2. RESULTS: CHATGPT AND STUDENTS' BELIEFS AND PREJUDICES ON PLAGIARISM

The first question about their beliefs was to assess their level of agreement with the statement “ChatGPT did my work in this task” on a scale from 1 to 10. The answers are displayed in Figure 5.

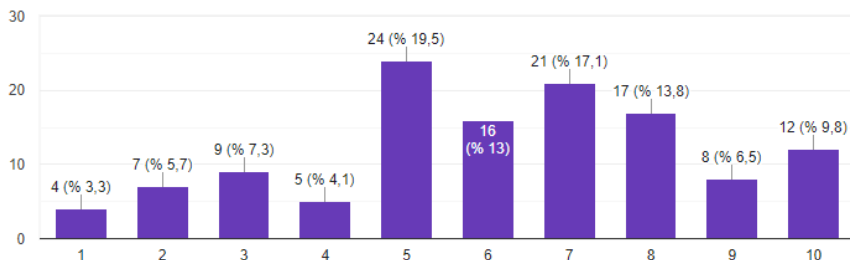
**FIGURE 5.** Degree of agreement with the assessment “ChatGPT did my work in this task”



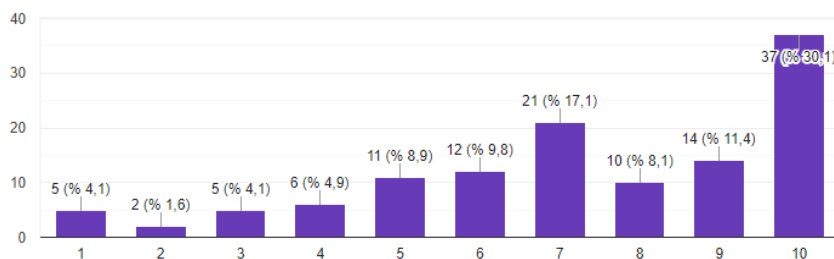
Figures 6 and 7 represent graphically the answers to the questions about the need of crediting ChatGPT as an author and their opinion about the same aspect if the work had been done by a person instead of by a tool. 19.4% of the students consider that the tool should not be

credited, which seems to indicate that they underestimate the support offered by the tool. Surprisingly enough, 14.7% of the students also answered that they do not think that the person should be credited provided that the same work had been carried out by a person.

**FIGURE 6.** Degree of agreement with the assessment “I should credit ChatGPT as an author”



**FIGURE 7.** Degree of agreement with the assessment “If the work done by Chatgpt were done by a person, I should credit Chatgpt as an author”



## 7. DISCUSSION

This section will comment on the results by drawing from the questionnaires and the experts’ evaluations.

On the quality of the proposals submitted by the students, the experts’ committee identified several issues. The overarching problem was, as Table 2 shows above, that students used ChatGPT without critical thinking, which the tool is incapable of, as seen in the literature section (Bishop, 2023). Students fed the chat a low number of commands,



resulting in impractical activities and disregard for the methodological principles of CLT and FC.

Although the texts produced did not have any language errors that does not mean that activities or the language level are appropriate. More specifically, the experts found that students did not take into account various aspects of CLT theory such as integration of skills (activities rarely integrated more than two skills) and lack of group dynamics. Moreover, the materials were not adapted to the language and cognitive level of primary students: the language in the video was too complex, explanations were too lengthy and cognitively demanding and activities were repetitive and sometimes too difficult for primary students.

Surprisingly, the layer of control activity (a crucial step in any FC lesson), which was the easiest task to prepare, (a labelling activity, a gap-filling activity, or identifying the key terms would be enough) gets the lowest score (a mean value of 6.42 in a scale 1-10) in terms of perceived usefulness. Paradoxically, after assessing these, the experts have identified that students may not have fully grasped this concept.

With regard to the design of the didactic proposals, there was a lack of integrative perspective, which revealed that the students did not consider the sequence globally. Among the most common pitfalls, there was a lack of sequencing insight, with a disorderly and non-progressive increase in cognitive and linguistic difficulty among activities. In that regard, the video and the activities were not connected. Additionally, some activities did not develop high-order thinking skills (HOTS) to apply the knowledge presented in the video and tasks were often limited to simple operations such as comprehension questions. Repeated activities were also found among different groups' submissions, further confirming the idea that some activities have been submitted without any or minimum modifications. And finally, a non-realistic measuring of timing was also observable in the description of the activities, students underestimating the time required for an activity.

Regarding assessment, the students did not design feasible evaluation for their proposals. In many cases, the assessment does not attain the goal of evaluating if the content has been learnt, partially because the tasks did not cover it correctly. Pedagogical aspects such as how to evaluate each exercise, offering support or calling reasoning to any thinking skill are considered neither by the tool nor by the students.

All the abovementioned problems that the experts' committee observed in the students' proposals contrast with the students' perceptions of the usefulness of the tool as illustrated in Figure 3. The numbers show a disparity in the perception of the students, overstating ChatGPT's capacity while failing to see structural problems in their designs. The latter could be explained by the following elements. First, the students are enrolled at an introductory level course in the didactics of foreign language and their lack of expertise in the field affects their criteria to judge how methodologically sound their proposals are. Second, the course is one of the few courses taught in English in the degree, the majority of students having a language level between B1 and B2 CEFR. It is plausible that students confuse ChatGPT's proficiency in English for methodological appropriateness, which would also explain both the high ratings that the tool received as far as its potential and the contrasting low number of commands the students fed the AI.

In contrast, students who modified the activities and materials provided by ChatGPT showed more practicality in their designs as per the opinions of the experts' committee. In that regard, the answers for subtasks ST4, ST5 and ST6 in table 1 are quite revealing among those students whose groups received the highest scores from the experts. The groups that registered a higher number of iterations received the highest scores.

Finally, there was the case of group no.25 which only used AI to design one activity out of the required three CLT activities. While these last specific results are inconclusive, it must be mentioned that the activities created without the tool's support show a higher degree of realism, practicality and understanding of the CLT principles.

There is a direct relationship between the number of iterations and the assessment of the experts. Thus, the case of this last group suggests that students who might have been able to develop interesting and correct proposals might have been hindered by the cost of opportunity of using the tool instead of their own capabilities. This does not contradict the fact that those groups who have fed the chat with a high count of commands have created appropriate proposals, showing practicality, language and cognitive adaptations to primary students, and consistency of sequence. Therefore, we conclude that with appropriate commands and specifying the language level, plus making follow-up questions to simplify and improve the design, didactic results with ChatGPT are possible and viable (McMurtrie, 2023, cited in Rudolph et al, 2023).

As far as how students perceive the tool, the responses to the open questions yielded contradictory and worrying results (particularly with regard to plagiarism and the potential unethical use of the AI). In terms of usefulness, the students valued two elements: the ability to generate responses in English, a considerable advantage since most of them are not proficient, and the fact that ChatGPT can provide exercises, adapt the length and complexity of the content and provide accurate knowledge (which has been questioned by recent studies, as seen above: Bishop, 2023; Bowman, 2022; ChatGPT & Perlman; Llorens-Largo & ChatGPT, 2022; Pearl, 2002).

On the issue of plagiarism, results reveal contradictory beliefs held by the students. The progressive increase in agreement when crediting ChatGPT or a human (figure 6 and figure 7) shows how at odds the perceptions of students are. The antithetical position is further stressed by the fact that most students agree that ChatGPT did the job (figure 5). This triangulation of the data seems to indicate how some students have a very inconsistent idea of what constitutes plagiarism. An additional contradictory tendency can be spotted when comparing said inconsistent understanding of what constitutes plagiarism from a theoretical perspective and what they consider about this particular practice. This is reinforced by the risks identified in the open questions: the increase in plagiarism, the lack of reliability of the

information provided by the AI and the impossibility to know which sources ChatGPT is using to the best of their knowledge.

In conclusion, even if they could identify the problems that a widespread use of the tool may create in higher education, their inconsistent beliefs on plagiarism and the contradictory tendencies when crediting an AI or a human pinpoint a general tendency to identify problems with plagiarism in a general context. However, there is a blindness when it comes to seeing how their attitudes are key elements in fostering the issues. In other words, the respondents showed an antithetical tendency between their perceptions of plagiarism and their practices, in line with what has been observed by pre-existing literature (Pittam et al, 2009; Riskey et al, 2013).

Unfortunately, there seem to be more reasons for concern. In addition to the perception held by some students that even if acknowledging that the AI did their job, they would not credit it, the fact that 14.7% consider that a person should not be credited in the same situation, leaves a worrying impression. Since no previous enquiries on the perceptions of plagiarism of the respondents were conducted, we need to account for ignorance as an explanation, among others, for this potential unethical behaviour on the part of these students. Regardless of the reason, the dangers are to be accounted for. Particularly, because the dishonest use of the AI seems not to constitute a problem among some students (figure 6) and especially since the majority have admitted they would use it for future university assessment tasks. Although the previous question did not specify whether they would use it for research or dishonest purposes, the abovementioned contradictory beliefs and inconsistent perceptions of what constitutes plagiarism among students serve as a basis for a worrying picture.

## 8. CONCLUSIONS

It can be concluded that it is evident that the tool has many limitations. While it serves as a good starting point and is useful for rudimentary tasks such as writing guidelines, it lacks practicality and realism when it comes to designing activities. However, ChatGPT has proven to be a

valuable source of ideas that can be developed into feasible activities, provided that users modify the results according to the cognitive and linguistic level of the targeted group, the methodological principles involved, and appropriate sequencing.

It is also apparent that students may not be familiar with how to use the tool and have not fully mastered the methodological principles of CLT and FC, so different results may be obtained provided that these conditions are modified. Therefore, a possibility for future research could be to focus on teaching students how to better utilise the tool to their advantage or for example, obtain the sources of the information, and experiment with students who have more experience in these methodologies.

Another worrying conclusion that arises from the research is the evidence of plagiarism and students' intentions or at least the lack of concern in some cases about using materials and ideas created by others, being the co-author a person or an application, and presenting them as one's own. It is crucial to consider a hands-on intervention to raise awareness of the issue of plagiarism.

In summary, while ChatGPT has limitations, it is still a valuable tool for generating ideas that should be adapted by applying the teacher's professional expertise and criteria to suit the particular needs of each situation. The results of the study suggest that further research is needed to enhance university students' understanding of how to utilise the tool effectively and to address the issue of plagiarism.

## 9. REFERENCES

Anderson, N., Belavy, D. L., Perle, S. M., Hendricks, S., Hespanhol, L., Verhagen, E., & Memon, A. R. (2023). AI did not write this manuscript, or did it? Can we trick the AI text detector into generated texts? The potential future of CHATGPT and AI in Sports & Exercise Medicine Manuscript Generation. *BMJ Open Sport & Exercise Medicine*, 9(1).  
<https://doi.org/10.1136/bmjsem-2023-00156>

Atlas, S. (2023). ChatGPT for Higher Education and Professional Development: A Guide to Conversational AI.

[https://digitalcommons.uri.edu/cba\\_facpubs/548](https://digitalcommons.uri.edu/cba_facpubs/548)

Baidoo-Anu, D. & Owusu Ansah, L. (2023) Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. Available at SSRN:

<https://ssrn.com/abstract=4337484> or  
<http://dx.doi.org/10.2139/ssrn.4337484>

Bishop, L. (2023) A Computer Wrote this Paper: What ChatGPT Means for Education, Research, and Writing. Available at SSRN:

<https://ssrn.com/abstract=4338981> or  
<http://dx.doi.org/10.2139/ssrn.4338981>

Bowman, E. (2022, December 19). A new AI chatbot might do your homework for you. But it's still not an A+ student. *NPR*. <http://bit.ly/3QL6z8A>

Brimble, M., & Stevenson-Clarke, P. (2005). Perceptions of the prevalence and seriousness of academic dishonesty in Australian universities. *Australian Educational Researcher*, 32, 19–44.

ChatGPT, & Perlman, A. (2022). The Implications of OpenAI's Assistant for Legal Services and Society. Available at SSRN:

<https://ssrn.com/abstract=4294197> or  
<http://dx.doi.org/10.2139/ssrn.4294197>

Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education: A Review. *IEEE Access*, 8, 75264-75278.

<https://doi.org/10.1109/ACCESS.2020.2988510>

Codina, L. (2022, December 12). Cómo utilizar ChatGPT en el aula con perspectiva ética y pensamiento crítico: Una proposición para docentes y educadores. <http://bit.ly/3iKBFAE>

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 1–12.

<https://doi.org/10.1080/14703297.2023.2190148>

Dehouche, N. (2021). Plagiarism in the age of massive generative pre-trained

Transformers (GPT-3). *Ethics in Science and Environmental Politics*, 21, 17–23. <https://doi.org/10.3354/esep00195>

Frye, B.L. (2022). Should using an AI text generator to produce academic writing be plagiarism? *Fordham Intellectual Property, Media & Entertainment Law Journal*, Forthcoming. Available at SSRN: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4292283](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4292283)

Gagné, Robert M. (1985). *The Conditions of Learning and Theory of Instruction* (4th Edition). CBS College Publishing.

Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2023). Comparing scientific abstracts generated by CHATGPT to real abstracts with detectors and blinded human reviewers. *Npj Digital Medicine*, 6(1). <https://doi.org/10.1038/s41746-023-00819-6>

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Howard, R. M. (1995). Plagiarisms, authorships, and the Academic Death Penalty. *College English*, 57(7), 788. <https://doi.org/10.2307/378403>

Howard, S., Enrich, J.F. & Walton, R. (2014). Measuring Students' Perceptions of Plagiarism: Modification and Rasch Validation of a Plagiarism Attitude Scale. *Journal of Applied Measurement*, 15(4), 372-393. [https://www.researchgate.net/publication/265791453\\_Measuring\\_Students'\\_Perceptions\\_of\\_Plagiarism\\_Modification\\_and\\_Rasch\\_Validation\\_of\\_a\\_Plagiarism\\_Attitude\\_Scale](https://www.researchgate.net/publication/265791453_Measuring_Students'_Perceptions_of_Plagiarism_Modification_and_Rasch_Validation_of_a_Plagiarism_Attitude_Scale)

Hu, G., & Lei, J. (2015). Chinese university students' perceptions of plagiarism. *Ethics & Behavior*, 25(3), 233–255. <https://doi.org/10.1080/10508422.2014.923313>

Llorens-Largo, F., & ChatGPT. (2022, December 22). Cavilaciones invernales. Universidad. <http://bit.ly/3XGk0Jn>

McCabe, D. L., & Bowers, W. J. (1994). Academic dishonesty among males in college: A thirty year perspective. *Journal of College Student Development*, 35, 5–10.

- Ouyang, F., & Jiao, P. (2021). Artificial Intelligence in Education: The Three Paradigms. *Computers and Education: Artificial Intelligence*, 2, 100020. <https://doi.org/10.1016/j.caeai.2021.100020>
- Osmanovic-Thunström, A., & Steingrímsson, S. (2023). Does GPT-3 qualify as a co-author of a scientific paper publishable in peer-review journals according to the ICMJE criteria? A case study. *Discover Artificial Intelligence*, 3(1). <https://doi.org/10.1007/s44163-023-00055-7>
- Pavlik, J. V. (2023). Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education. *Journalism & Mass Communication Educator*, 78(1), 84–93. <https://doi.org/10.1177/10776958221149577>
- Pearl, M. (2022, December 3). The ChatGPT chatbot from OpenAI is amazing, creative, and totally wrong. Need ideas? Great! Need facts? Stay away! Mashable. <http://bit.ly/3ZJAkec>
- Pickell, T. R., & Doak, B. R. (2023). Five Ideas for How Professors Can Deal with GPT-3 ... For Now. *Faculty Publications - George Fox School of Theology*. 432. <https://digitalcommons.georgefox.edu/ccs/432>
- Pittam, G., Elander, J., Lusher, J., Fox, P., & Payne, N. (2009). Student beliefs and attitudes about authorial identity in academic writing. *Studies in Higher Education*, 34(2), 153–170. <https://doi.org/10.1080/03075070802528270>
- Pressey, S.L. (1950). Development and appraisal of devices providing immediate automatic scoring of objective tests and concomitant self-instruction. *Journal of Psychology* 30, 417–447.
- Rudolph, J., Tan, S. & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching* 6(1), 1-22.
- Risque, A., O'Dwyer, M., & Ledwith, A. (2013). 'thou shalt not plagiarise': From self-reported views to recognition and avoidance of plagiarism. *Assessment & Evaluation in Higher Education*, 38(1), 34–43. <https://doi.org/10.1080/02602938.2011.596926>



- Srinivasan, V. & Murthy, H. (2021). Improving reading and comprehension in K-12: Evidence from a large-scale AI technology intervention in India. *Computers and Education: Artificial Intelligence* 2, 100019. <https://doi.org/10.1016/j.caeai.2021.100019>
- Skinner, B.F. (1958). Teaching machines. *Science* 128 (3330), 969-977.
- Stokel-Walker, C. (2022). Ai Bot chatgpt writes smart essays — should professors worry? *Nature*. <https://doi.org/10.1038/d41586-022-04397-7>
- Tahiru, F. (2021). AI in education: A systematic literature review. *Journal of Cases on Information Technology* 23(1), 1-20.
- Vincent-Lancrin, S. & van der Vlies, R.. (2020). Trustworthy artificial intelligence (AI) in education: Promises and challenges. *OECD Education Working Papers*, 218. <https://doi.org/10.1787/a6c90fa9-en>
- Vygotsky, L. C. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Williamson, B. & Eynon, R. (2020) Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology* 45(3), 223-235. <https://doi.org/10.1080/17439884.2020.1798995>